SIAM J. IMAGING SCIENCES
Vol. xx, pp. x x–x

(IN PRESS[a])

Description of the minimizers of least squares regularized with ℓ_0-norm. Uniqueness of the global minimizer

Mila Nikolova*

Abstract. We have an $\mathsf{M} \times \mathsf{N}$ real-valued arbitrary matrix A (e.g. a dictionary) with $\mathsf{M} < \mathsf{N}$ and data d describing the sought-after object with the help of A. This work provides an in-depth analysis of the (local and global) minimizers of an objective function $\mathcal{F}_d$ combining a quadratic data-fidelity term and an ℓ_0 penalty applied to each entry of the sought-after solution, weighted by a regularization parameter $\beta > 0$. For several decades, this objective has attracted a ceaseless effort to conceive algorithms approaching a good minimizer. Our theoretical contributions, summarized below, shed new light on the existing algorithms and can help the conception of innovative numerical schemes. To solve the normal equation associated with any M-row submatrix of A is equivalent to compute a local minimizer $\hat{u}$ of $\mathcal{F}_d$. (Local) minimizers $\hat{u}$ of $\mathcal{F}_d$ are strict if and only if the submatrix, composed of those columns of A whose indexes form the support of $\hat{u}$, has full column rank. An outcome is that strict local minimizers of $\mathcal{F}_d$ are easily computed without knowing the value of β. Each strict local minimizer is linear in data. It is proved that $\mathcal{F}_d$ has global minimizers and that they are always strict. They are studied in more details under the (standard) assumption that $\mathrm{rank}(A) = \mathsf{M} < \mathsf{N}$. The global minimizers with M-length support are seen to be impractical. Given d, critical values β_{K} for any $\mathsf{K} \leqslant \mathsf{M} - 1$ are exhibited such that if $\beta > \beta_{\mathsf{K}}$, all global minimizers of $\mathcal{F}_d$ are K-sparse. An assumption on A is adopted and proved to fail only on a closed negligible subset. Then for all data d beyond a closed negligible subset, the objective $\mathcal{F}_d$ for $\beta > \beta_{\mathsf{K}}$, $\mathsf{K} \leqslant \mathsf{M} - 1$, has a unique global minimizer and this minimizer is K-sparse. Instructive small-size (5×10) numerical illustrations confirm the main theoretical results.

Key words. asymptotically level stable functions, global minimizers, local minimizers, ℓ_0 regularization, nonconvex nonsmooth minimization, perturbation analysis, quadratic programming, solution analysis, sparse recovery, strict minimizers, underdetermined linear systems, uniqueness of the solution, variational methods

AMS subject classifications. 15A03, 15A29, 15A99, 26E35, 28A75, 46N99, 47J07, 49J99, 49K40, 90C26, 90C27, 94A08, 94A12.

DOI. 10.1137/11085476X

1. Introduction.

Let A be an arbitrary matrix (e.g., a dictionary) such that

$$A \in \mathbb{R}^{\mathsf{M} \times \mathsf{N}} \quad \text{for} \quad \mathsf{M} < \mathsf{N} \ ,$$

where the positive integers M and N are fixed. Given a data vector $d \in \mathbb{R}^{\mathsf{M}}$, we consider an objective function $\mathcal{F}_d : \mathbb{R}^{\mathsf{N}} \to \mathbb{R}$ of the form

(1)
$$\mathcal{F}_d(u) = \|Au - d\|_2^2 + \beta \|u\|_0 \ , \quad \beta > 0 \ ,$$
$$\|u\|_0 = \sharp\, \sigma(u) \ ,$$

*CMLA, ENS Cachan, CNRS, 61 Av. Prés. Wilson, 94230 Cachan, France (nikolova@cmla.ens-cachan.fr).

[a]Received by the editors November 10, 2011; accepted for publication (in revised form) January 14, 2013; published electronically DATE.

where $u \in \mathbb{R}^{\mathsf{N}}$ contains the coefficients describing the sought-after object, β is a regularization parameter, $\sharp$ stands for cardinality and $\sigma(u)$ is the support of u (i.e., the set of all $i \in \{1, \cdots, \mathsf{N}\}$ for which the ith entry of u satisfies $u[i] \neq 0$). By an abuse of language, the penalty in (1) is called the ℓ_0-norm. Define $\phi : \mathbb{R} \to \mathbb{R}$ by

$$\phi(t) \stackrel{\text{def}}{=} \begin{cases} 0 & \text{if} \quad t = 0 \, , \\ 1 & \text{if} \quad t \neq 0 \, . \end{cases} \tag{2}$$

Then $\|u\|_0 = \sum_{i=1}^{\mathsf{N}} \phi(u[i]) = \sum_{i \in \sigma(u)} \phi(u[i])$, so $\mathcal{F}_d$ in (1) equivalently reads

$$\mathcal{F}_d(u) = \|Au - d\|_2^2 + \beta \sum_{i=1}^{\mathsf{N}} \phi(u[i]) = \|Au - d\|_2^2 + \beta \sum_{i \in \sigma(u)} \phi(u[i]) \ . \tag{3}$$

We focus on all (local and global) minimizers $\hat{u}$ of an objective $\mathcal{F}_d$ of the form (1):

$$\hat{u} \in \mathbb{R}^{\mathsf{N}} \ \text{ such that } \ \mathcal{F}_d(\hat{u}) = \min_{u \in \mathcal{O}} \mathcal{F}_d(u) \ , \tag{4}$$

where $\mathcal{O}$ is an open neighborhood of $\hat{u}$. We note that finding a global minimizer of $\mathcal{F}_d$ must be an *NP-hard* computational problem [11, 40].

The function ϕ in (2) served as a regularizer for a long time. In the context of Markov random fields it was used by Geman and Geman in 1984 [20] and Besag in 1986 [5] as a prior in MAP energies to restore labeled images. The MAP objective reads as

$$\mathcal{F}_d(u) = \|Au - d\|_2^2 + \beta \sum_k \phi(D_k u) \ , \tag{5}$$

where D_k is a finite difference operator and ϕ is given by (2). This label-designed form is known as the Potts prior model, or as the multi-level logistic model [6, 24]. Various stochastic and deterministic algorithms have been considered to minimize (5). Leclerc [23] proposed in 1989 a deterministic continuation method to restore piecewise constant images. Robini, Lachal and Magnin [33] introduced the stochastic continuation approach and successfully used it to reconstruct 3D tomographic images. Robini and Magnin refined the method and the theory in [34]. Very recently, Robini and Reissman [35] gave theoretical results relating the probability for global convergence and the computation speed.

The problem stated in (1) and (4)—to (locally) minimize $\mathcal{F}_d$—arises when *sparse* solutions are desired. Typical application fields are signal and image processing, morphologic component analysis, compression, dictionary building, inverse problems, compressive sensing, machine learning, model selection, classification, and subset selection, among others. The original hard-thresholding method proposed by Donoho and Johnstone [15] amounts to[1] minimizing $\mathcal{F}_d$, where d contains the coefficients of a signal or an image expanded in a wavelet basis ($\mathsf{M} = \mathsf{N}$). When $\mathsf{M} < \mathsf{N}$, various (usually strong) restrictions on $\|u\|_0$ (often $\|u\|_0$ is replaced by a less

[1] As a reminder, if d are some noisy coefficients, the restored coefficients $\hat{u}$ minimize $\|u - d\|^2 + \beta \|u\|_0$ and read $\hat{u}[i] = 0$ if $|d[i]| \leqslant \sqrt{\beta}$ and $\hat{u}[i] = d[i]$ otherwise.

irregular function) and on A (e.g., RIP-like criteria, conditions on $\|A\|$, etc.) are needed to conceive numerical schemes approximating a minimizer of $\mathcal{F}_d$, to establish local convergence and derive the asymptotic of the obtained solution. In statistics the problem has been widely considered for subset selection, and numerous algorithms have been designed, with limited theoretical production, as explained in the book by Miller [30]. More recently, Haupt and Nowak [22] investigate the statistical performances of the global minimizer of $\mathcal{F}_d$ and propose an iterative bound-optimization procedure. Fan and Li [17] discuss a variable splitting and penalty decomposition minimization technique for (1), along with other approximations of the ℓ_0-norm. Liu and Wu [25] mix the ℓ_0 and ℓ_1 penalties, establish some asymptotic properties of the new estimator and use mixed integer programming aimed at global minimization. For model selection, Lv and Fan [27] approximate the ℓ_0 penalty using functions that are concave on $\mathbb{R}_+$ and prove a nonasymptotic nearly oracle property of the resultant estimator. Thiao, Dinh, and Thi [39] reformulate the problem so that an approximate solution can be found using difference-of-convex-functions programming. Blumensath and Davies [7] propose an iterative thresholding scheme to approximate a solution and prove convergence to a local minimizer of $\mathcal{F}_d$. Lu and Zhang [26] suggest a penalty decomposition method to minimize $\mathcal{F}_d$. Fornasier and Ward [18] propose an iterative thresholding algorithm for minimizing $\mathcal{F}_d$ where ℓ_0 is replaced by a reasonable sparsity-promoting relaxation given by $\phi(t) = \min\{|t|, 1\}$; then convergence to a local minimizer is established. In a recent paper by Chouzenoux et al. [9], a mixed $\ell_2 - \ell_0$ regularization is considered: a slightly smoothed version of the objective is analyzed and a majorize-minimize subspace approach, satisfying a finite length property, converges to a critical point. Since the submission of our paper, image reconstruction methods have been designed where ℓ_0 regularization is applied to the coefficients of the expansion of the sought-after image in a wavelet frame [42, 14]: the provided numerical results outperform ℓ_1 regularization for a reasonable computational cost achieved using penalty decomposition techniques. In a general study on the convergence of descent methods for nonconvex objectives, Attouch, Bolte, and Svaiter [1] apply an inexact forward-backward splitting scheme to find a critical point of $\mathcal{F}_d$. Several other references can be evoked, e.g., [31, 19].

Even though overlooked for several decades, the objective $\mathcal{F}_d$ was essentially considered from a numerical standpoint. The motivation naturally comes from the promising applications and the intrinsic difficulty of minimizing $\mathcal{F}_d$.

The goal of this work is to analyze the (local and global) minimizers $\hat{u}$ of objectives $\mathcal{F}_d$ of the form (1)*.*

- *We provide detailed results on the minimization problem.*
- *The uniqueness of the global minimizer of $\mathcal{F}_d$ is examined as well.*

We do not propose an algorithm. However, our theoretical results raise salient questions about the existing algorithms and can help the conception of innovative numerical schemes.

The minimization of $\mathcal{F}_d$ in (1) might seem close to its constraint variants:

$$\begin{array}{llll} \text{given } \varepsilon \geqslant 0, & \text{minimize } \ \|u\|_0 & \text{subject to } \ \|Au-d\|^2 \leqslant \varepsilon \ , \\ \text{given } K \in \mathbb{I}_{\mathsf{M}}, & \text{minimize } \|Au-d\|^2 & \text{subject to } \ \|u\|_0 \leqslant K \ . \end{array} \tag{6}$$

The latter problems are abundantly studied in the context of sparse recovery in different fields. An excellent account is given in [8], see also the book [28]. For recent achievements, we refer the reader to [10]. It is worth emphasizing that in general, *there is no equivalence between the*

problems stated in (6) *and the minimization of* $\mathcal{F}_d$ *in* (1) because all of these problems are nonconvex.

1.1. Main notation and definitions. We recall that if $\hat{u}$ is a *(local) minimizer* of $\mathcal{F}_d$, the value $\mathcal{F}_d(\hat{u})$ is a (local) minimum[2] of $\mathcal{F}_d$ reached at (possibly numerous) points $\hat{u}$. Saying that a (local) minimizer $\hat{u}$ is *strict* means that there is a neighborhood $\mathcal{O} \subset \mathbb{R}^{\mathsf{N}}$, containing $\hat{u}$, such that $\mathcal{F}_d(\hat{u}) < \mathcal{F}_d(v)$ for any $v \in \mathcal{O} \setminus \{\hat{u}\}$. So $\hat{u}$ is an isolated minimizer.

Let K be any positive integer. The expression $\{u \in \mathbb{R}^{\mathsf{K}} \ : \ u \ \text{ satisfying property } \ \mathfrak{P}\}$ designates the subset of $\mathbb{R}^{\mathsf{K}}$ formed from all elements u that meet $\mathfrak{P}$. The identity operator on $\mathbb{R}^{\mathsf{K}}$ is denoted by I_{K}. The entries of a vector $u \in \mathbb{R}^{\mathsf{K}}$ read as $u[i]$, for any i. The ith vector of the canonical basis[3] of $\mathbb{R}^{\mathsf{K}}$ is denoted by $e_i \in \mathbb{R}^{\mathsf{K}}$. Given $u \in \mathbb{R}^{\mathsf{K}}$ and $\rho > 0$, the *open* ball at u of radius ρ with respect to the ℓ_p-norm for $1 \leqslant p \leqslant \infty$ reads as

$$\mathrm{B}_p(u,\rho) \stackrel{\text{def}}{=} \{v \in \mathbb{R}^{\mathsf{K}} \ : \ \|v-u\|_p < \rho\} \ .$$

To simplify the notation, the ℓ_2-norm is *systematically* denoted by

$$\|\cdot\| \stackrel{\text{def}}{=} \|\cdot\|_2 \ .$$

We denote by $\mathbb{I}_{\mathsf{K}}$ the *totally and strictly ordered* index set[4]

$$\mathbb{I}_{\mathsf{K}} \stackrel{\text{def}}{=} \big(\{1,\cdots,\mathsf{K}\}, < \big) \ , \tag{7}$$

where the symbol $<$ stands for the natural order of the positive integers. Accordingly, *any subset* $\omega \subseteq \mathbb{I}_{\mathsf{K}}$ *inherits the property of being totally and strictly ordered.*

We shall often consider the index set $\mathbb{I}_{\mathsf{N}}$. The complement of $\omega \subseteq \mathbb{I}_{\mathsf{N}}$ in $\mathbb{I}_{\mathsf{N}}$ is denoted by

$$\omega^c = \mathbb{I}_{\mathsf{N}} \setminus \omega \subseteq \mathbb{I}_{\mathsf{N}} \ .$$

Definition 1.1. *For any* $u \in \mathbb{R}^{\mathsf{N}}$, *the* support $\sigma(u)$ *of* u *is defined by*

$$\sigma(u) = \Big\{ i \in \mathbb{I}_{\mathsf{N}} \ : \ u[i] \neq 0 \Big\} \subseteq \mathbb{I}_{\mathsf{N}} \ .$$

If $u = 0$, clearly $\sigma(u) = \varnothing$.

The ith column in a matrix $A \in \mathbb{R}^{\mathsf{M}\times\mathsf{N}}$ is denoted by a_i. It is *systematically* assumed that

$$\boxed{a_i \neq 0 \quad \forall\, i \in \mathbb{I}_{\mathsf{N}} \ .} \tag{8}$$

For a matrix $A \in \mathbb{R}^{\mathsf{M}\times\mathsf{N}}$ and a vector $u \in \mathbb{R}^{\mathsf{N}}$, with any $\omega \subseteq \mathbb{I}_{\mathsf{N}}$, we associate the *submatrix* A_ω and the *subvector* u_ω given by

$$A_\omega \stackrel{\text{def}}{=} \big(a_{\omega[1]}, \cdots, a_{\omega[\sharp\omega]}\big) \in \mathbb{R}^{\mathsf{M}\times\sharp\omega} \ , \tag{9}$$

$$u_\omega \stackrel{\text{def}}{=} \Big(u\big[\omega[1]\big], \cdots, u\big[\omega[\sharp\omega]\big]\Big) \in \mathbb{R}^{\sharp\omega} \ , \tag{10}$$

---

[2] These two terms are often confused in the literature.

[3] More precisely, for any $i \in \mathbb{I}_{\mathsf{K}}$, the vector $e_i \in \mathbb{R}^{\mathsf{K}}$ is defined by $e_i[i] = 1$ and $e_i[j] = 0$, $\forall\, j \in \mathbb{I}_{\mathsf{K}} \setminus \{i\}$.

[4] E.g. without strict order we have $\omega = \{1,2,3\} = \{2,1,1,3\}$ in which case the notation in (9)-(10) below is ambiguous.

respectively, as well as the zero padding operator $Z_\omega \;:\; \mathbb{R}^{\sharp\omega} \to \mathbb{R}^{\mathsf{N}}$ that inverts (10):

$$u = Z_\omega\left(u_\omega\right) \;,\quad u[i] = \begin{cases} 0 & \text{if } \; i \notin \omega \;, \\ u_\omega[k] & \text{for the unique } k \text{ such that } \; \omega[k] = i. \end{cases} \tag{11}$$

Thus for $\omega = \varnothing$ one finds $u_\varnothing = \varnothing$ and $u = Z_\varnothing\left(u_\varnothing\right) = 0 \in \mathbb{R}^{\mathsf{N}}$.

Using Definition 1.1 and the notation in (9)-(10), for any $u \in \mathbb{R}^{\mathsf{N}} \setminus \{0\}$ we have

$$\omega \in \mathbb{I}_{\mathsf{N}} \;\text{ and }\; \omega \supseteq \sigma(u) \quad \Rightarrow \quad Au = A_\omega u_\omega \;. \tag{12}$$

To simplify the presentation, we adopt the following *definitions* [5]:

$$\begin{array}{ll} (a) & A_\varnothing = [\;] \in \mathbb{R}^{\mathsf{M}\times 0} \;, \\ (b) & \operatorname{rank}\left(A_\varnothing\right) = 0 \;. \end{array} \tag{13}$$

In order to avoid possible ambiguities[6], we set

$$A_\omega^T \stackrel{\text{def}}{=} \left(A_\omega\right)^T \;,$$

where the superscript ^{T} stands for transposed. If A_ω is invertible, similarly $A_\omega^{-1} \stackrel{\text{def}}{=} \left(A_\omega\right)^{-1}$.

In the course of this work, we shall frequently refer to the constrained quadratic optimization problem stated next.

Given $d \in \mathbb{R}^{\mathsf{M}}$ and $\omega \subseteq \mathbb{I}_{\mathsf{N}}$, problem $(\,\mathcal{P}_\omega\,)$ reads as:

$$\boxed{\begin{cases} & \min\limits_{u \in \mathbb{R}^{\mathsf{N}}} \|Au - d\|^2 \;, \\ \text{subject to} & u[i] = 0, \quad \forall\, i \in \omega^c \;. \end{cases}} \qquad (\,\mathcal{P}_\omega\,) \tag{14}$$

Clearly, problem $(\,\mathcal{P}_\omega\,)$ always admits a solution.

The definition below will be used to evaluate the extent of some subsets and assumptions.

Definition 1.2. *A property (an assumption) is called* generic on $\mathbb{R}^{\mathsf{K}}$ *if it holds true on a* dense open *subset of* $\mathbb{R}^{\mathsf{K}}$.

As usual, a subset $\mathcal{S} \subset \mathbb{R}^{\mathsf{K}}$ is said to be *negligible* in $\mathbb{R}^{\mathsf{K}}$ if there exists $\mathcal{Z} \subset \mathbb{R}^{\mathsf{K}}$ whose Lebesgue measure in $\mathbb{R}^{\mathsf{K}}$ is $\mathbb{L}^{\mathsf{K}}(\mathcal{Z}) = 0$ and $\mathcal{S} \subseteq \mathcal{Z}$. If a property fails only on a negligible set, it is said to hold *almost everywhere*, meaning "with probability one". Definition 1.2 requires much more than *almost everywhere*. Let us explain.

If a property holds true for all $v \in \mathbb{R}^{\mathsf{K}} \setminus \mathcal{S}$, where $\mathcal{S} \subseteq \mathcal{Z} \subset \mathbb{R}^{\mathsf{K}}$, $\mathcal{Z}$ is closed in $\mathbb{R}^{\mathsf{K}}$ *and $\mathbb{L}^{\mathsf{K}}(\mathcal{Z}) = 0$, then this property is* generic *on* $\mathbb{R}^{\mathsf{K}}$. Indeed, $\mathbb{R}^{\mathsf{K}} \setminus \mathcal{Z}$ contains a *dense open* subset of $\mathbb{R}^{\mathsf{K}}$. So if a property is generic on $\mathbb{R}^{\mathsf{K}}$, then it holds true almost everywhere on $\mathbb{R}^{\mathsf{K}}$. But the converse is false: an almost everywhere true property is not generic if the closure of its negligible subset has a positive measure,[7] because then $\mathbb{R}^{\mathsf{K}} \setminus \mathcal{Z}$ does not contains an open

---

[5] Note that (a) corresponds to the zero mapping on $\mathbb{R}^0$ and that (b) is the usual definition for the rank of an empty matrix.

[6] In the light of (9), A_ω^T could also mean $\left(A^T\right)_\omega$.

[7] There are many examples—e.g. $\mathcal{Z} = \{x \in [0,1] \;:\; x \text{ is rational}\}$, then $\mathbb{L}^1(\mathcal{Z}) = 0$ and $\mathbb{L}^1(\text{closure}(\mathcal{Z})) = 1$.

subset of $\mathbb{R}^{\mathsf{K}}$. In this sense, a generic property is stable with respect to the objects to which it applies.

The elements of a set $\mathcal{S} \subset \mathbb{R}^{\mathsf{K}}$ where a generic property fails are highly exceptional in $\mathbb{R}^{\mathsf{K}}$. The chance that a truly random $v \in \mathbb{R}^{\mathsf{K}}$—i.e., a v following a non singular probability distribution on $\mathbb{R}^{\mathsf{K}}$—comes across such an $\mathcal{S}$ can be ignored in practice.

1.2. Content of the paper. The main result in section 2 tells us that finding a solution of $(\mathcal{P}_\omega)$ for $\omega \subset \mathbb{I}_{\mathsf{N}}$ is *equivalent* to computing a (local) minimizer of $\mathcal{F}_d$. In section 3 we prove that the (local) minimizers $\hat{u}$ of $\mathcal{F}_d$ are *strict* if and only if the submatrix $A_{\sigma(\hat{u})}$ has full column rank. The strict minimizers of $\mathcal{F}_d$ are shown to be linear in data d. The importance of the $(\mathsf{M}-1)$-sparse strict minimizers is emphasized. The global minimizers of $\mathcal{F}_d$ are studied in section 4. Their existence is proved. They are shown to be strict for any d and for any $\beta > 0$. More details are provided under the standard assumption that $\mathrm{rank}(A) = \mathsf{M} < \mathsf{N}$. Given $d \in \mathbb{R}^{\mathsf{M}}$, critical values β_{K} for $\mathsf{K} \in \mathbb{I}_{\mathsf{M}-1}$ are exhibited such that all global minimizers of $\mathcal{F}_d$ are K-sparse[8] if $\beta > \beta_{\mathsf{K}}$.

In section 5, a gentle assumption on A is shown to be *generic* for all $\mathsf{M} \times \mathsf{N}$ real matrices. Under this assumption, for all data $d \in \mathbb{R}^{\mathsf{M}}$ beyond a closed negligible subset, the objective $\mathcal{F}_d$ for $\beta > \beta_{\mathsf{K}}$, $\mathsf{K} \in \mathbb{I}_{\mathsf{M}-1}$, has a unique global minimizer and this minimizer is K-sparse.

Small size (A is 5×10) numerical tests in section 6 illustrate the main theoretical results.

2. All minimizers of $\mathcal{F}_d$.

2.1. Preliminary results. First, we give some basic facts on problem $(\mathcal{P}_\omega)$ as defined in (14) that are needed for later use. If $\omega = \varnothing$, then $\omega^c = \mathbb{I}_{\mathsf{N}}$, so the unique solution of $(\mathcal{P}_\omega)$ is $\hat{u} = 0$. For an arbitrary $\omega \subset \mathbb{I}_{\mathsf{N}}$ meeting $\sharp\,\omega \geqslant 1$, $(\mathcal{P}_\omega)$ amounts to minimizing a quadratic term with respect to only $\sharp\,\omega$ components of u, the remaining entries being null. This *quadratic* problem $(\mathcal{Q}_\omega)$ reads as

$$\min_{v \in \mathbb{R}^{\sharp\,\omega}} \big\|A_\omega v - d\big\|^2, \quad \sharp\,\omega \geqslant 1\,, \qquad (\mathcal{Q}_\omega) \tag{15}$$

and it always admits a solution. Using the zero-padding operator Z_ω in (11), we have

$$\Big[\ \hat{u}_\omega \in \mathbb{R}^{\sharp\,\omega} \text{ solves } (\mathcal{Q}_\omega) \text{ and } \hat{u} = Z_\omega(\hat{u}_\omega)\ \Big] \quad \Leftrightarrow \quad \Big[\ \hat{u} \in \mathbb{R}^{\mathsf{N}} \text{ solves } (\mathcal{P}_\omega)\,, \quad \sharp\,\omega \geqslant 1\ \Big]\,.$$

The optimality conditions for $(\mathcal{Q}_\omega)$, combined with the definition in (13)(a), give rise to the following equivalence, which holds true for any $\omega \subseteq \mathbb{I}_{\mathsf{N}}$:

$$\Big[\ \hat{u} \in \mathbb{R}^{\mathsf{N}} \text{ solves } (\mathcal{P}_\omega)\ \Big] \quad \Leftrightarrow \quad \Big[\ \hat{u}_\omega \in \mathbb{R}^{\sharp\,\omega} \ \text{ solves } \ A_\omega^T A_\omega v = A_\omega^T d \ \text{ and } \ \hat{u} = Z_\omega(\hat{u}_\omega)\ \Big]. \tag{16}$$

Note that $A_\omega^T A_\omega v = A_\omega^T d$ in (16) is the normal equation associated with $A_\omega v = d$. The remark below shows that the optimal value of $(\mathcal{P}_\omega)$ in (14) can also be seen as an orthogonal projection problem.

---

[8] As usual, a vector u is said to be K-sparse if $\|u\|_0 \leqslant \mathsf{K}$.

Remark 1. Let $r \stackrel{\text{def}}{=} \text{rank}(A_\omega)$ and $B_\omega \in \mathbb{R}^{\mathsf{M}\times r}$ be an orthonormal basis for range(A_ω). Then $A_\omega = B_\omega H_\omega$ for a uniquely defined matrix $H_\omega \in \mathbb{R}^{r\times \sharp\omega}$ with rank$(H_\omega) = r$. Using (16), we have

$$A_\omega^T A_\omega \hat{u}_\omega = A_\omega^T d \quad \Leftrightarrow \quad {H_\omega}^T H_\omega \hat{u}_\omega = {H_\omega}^T {B_\omega}^T d \quad \Leftrightarrow \quad H_\omega \hat{u}_\omega = {B_\omega}^T d \quad \Leftrightarrow \quad A_\omega \hat{u}_\omega = B_\omega {B_\omega}^T d \;.$$

In addition, $\Pi_{\text{range}(A_\omega)} = B_\omega {B_\omega}^T$ is the orthogonal projector onto the subspace spanned by the columns of A_ω, see e.g. [29]. The expression above combined with (16) shows that

$$\Big[\, \hat{u} \in \mathbb{R}^{\mathsf{N}} \text{ solves } (\mathcal{P}_\omega) \Big] \quad \Leftrightarrow \quad \Big[\, \hat{u}_\omega \in \mathbb{R}^{\sharp\omega} \;\; \text{meets} \;\; A_\omega \hat{u}_\omega = \Pi_{\text{range}(A_\omega)} d \quad \text{and} \quad \hat{u} = Z_\varpi(\hat{u}_\varpi) \Big].$$

Obviously, $A\hat{u} = A_\omega \hat{u}_\omega$ is the orthogonal projection of d onto the basis B_ω.

For $\omega \subseteq \mathbb{I}_{\mathsf{N}}$, let K_ω denote the vector subspace

$$\mathrm{K}_\omega \stackrel{\text{def}}{=} \left\{ v \in \mathbb{R}^{\mathsf{N}} \;:\; v[i] = 0, \;\forall\, i \in \omega^c \right\} \;. \tag{17}$$

This notation enables problem $(\,\mathcal{P}_\omega\,)$ in (14) to be rewritten as

$$\min_{u \in \mathrm{K}_\omega} \|Au - d\|^2 \;. \tag{18}$$

The technical lemma below will be used in what follows. We emphasize that its statement is *independent* of the vector $\hat{u} \in \mathbb{R}^{\mathsf{N}} \setminus \{0\}$.

Lemma 2.1. *Let $d \in \mathbb{R}^{\mathsf{M}}$, $\beta > 0$, and $\hat{u} \in \mathbb{R}^{\mathsf{N}} \setminus \{0\}$ be* arbitrary*. For $\hat{\sigma} \stackrel{\text{def}}{=} \sigma(\hat{u})$, set*

$$\rho \stackrel{\text{def}}{=} \min \left\{ \min_{i \in \hat{\sigma}} \big|\, \hat{u}[i] \,\big|, \; \frac{\beta}{2\Big(\|A^T(A\hat{u} - d)\|_1 + 1\Big)} \right\}. \tag{19}$$

Then $\rho > 0$.

(i) *For ϕ as defined in (2), we have*

$$v \in \mathrm{B}_\infty(0,\rho) \quad \Rightarrow \quad \sum_{i \in \mathbb{I}_{\mathsf{N}}} \phi\big(\, \hat{u}[i] + v[i] \,\big) = \sum_{i \in \hat{\sigma}} \phi\,(\hat{u}[i]) + \sum_{i \in \hat{\sigma}^c} \phi\,(v[i]) \;\;.$$

(ii) *For $\mathrm{K}_{\hat{\sigma}}$ defined according to (17), $\mathcal{F}_d$ satisfies*

$$v \in \mathrm{B}_\infty(0,\rho) \cap \big(\mathbb{R}^{\mathsf{N}} \setminus \mathrm{K}_{\hat{\sigma}}\big) \quad \Rightarrow \quad \mathcal{F}_d(\hat{u} + v) \geqslant \mathcal{F}_d(\hat{u}) \;,$$

where the inequality is strict *whenever $\hat{\sigma}^c \neq \varnothing$.*

The proof is outlined in Appendix 8.1.

2.2. The (local) minimizers of $\mathcal{F}_d$ solve quadratic problems.

It is worth emphasizing that no special assumptions on the matrix A are adopted.

We begin with an easy but cautionary result.

Lemma 2.2. *For any $d \in \mathbb{R}^{\mathsf{M}}$ and for all $\beta > 0$, $\mathcal{F}_d$ has a* strict *(local) minimum at $\hat{u} = 0 \in \mathbb{R}^{\mathsf{N}}$.*

Proof. Using the fact that $\mathcal{F}_d(0) = \|d\|^2 \geqslant 0$, we have

$$\mathcal{F}_d(v) = \|Av - d\|^2 + \beta\|v\|_0 = \mathcal{F}_d(0) + \mathcal{R}_d(v) \ , \tag{20}$$

$$\text{where} \quad \mathcal{R}_d(v) = \|Av\|^2 - 2\langle v, A^T d\rangle + \beta\|v\|_0 \ . \tag{21}$$

Noticing that $\beta\|v\|_0 \geqslant \beta > 0$ for $v \neq 0$ leads to

$$v \in \mathrm{B}_2\left(0, \frac{\beta}{2\|A^T d\| + 1}\right) \setminus \{0\} \quad \Rightarrow \quad \mathcal{R}_d(v) \geqslant -2\|v\| \, \|A^T d\| + \beta > 0 \ .$$

Inserting this implication into (20) proves the lemma. □

For any $\beta > 0$ and $d \in \mathbb{R}^{\mathsf{M}}$, the sparsest strict local minimizer of $\mathcal{F}_d$ reads $\hat{u} = 0$. Initialization with zero of a suboptimal algorithm should generally be a bad choice. Indeed, experiments have shown that such an initialization can be harmful; see, e.g., [30, 7].

The next proposition states a result that is often evoked in this work.

Proposition 2.3. *Let $d \in \mathbb{R}^{\mathsf{M}}$. Given an $\omega \subseteq \mathbb{I}_{\mathsf{N}}$, let $\hat{u}$ solve problem $(\mathcal{P}_\omega)$ as formulated in (14). Then for any $\beta > 0$, the objective $\mathcal{F}_d$ in (1) reaches a (local) minimum at $\hat{u}$ and*

$$\sigma(\hat{u}) \subseteq \omega \ , \tag{22}$$

where $\sigma(\hat{u})$ is given in Definition 1.1.

Proof. Let $\hat{u}$ solve problem $(\mathcal{P}_\omega)$, and let $\beta > 0$. The constraint in $(\mathcal{P}_\omega)$ entails (22).

Consider that $\hat{u} \neq 0$, in which case for $\hat{\sigma} \overset{\text{def}}{=} \sigma(\hat{u})$ we have $1 \leqslant \sharp\hat{\sigma} \leqslant \sharp\omega$. Using the equivalent formulation of $(\mathcal{P}_\omega)$ given in (17)-(18), yields

$$v \in \mathrm{K}_\omega \quad \Rightarrow \quad \|A(\hat{u} + v) - d\|^2 \geqslant \|A\hat{u} - d\|^2 \ . \tag{23}$$

The inclusion in (22) is equivalent to $\omega^c \subseteq \hat{\sigma}^c$. Let $\mathrm{K}_{\hat{\sigma}}$ be defined according to (17) as well. Then

$$\hat{u} \in \mathrm{K}_{\hat{\sigma}} \subseteq \mathrm{K}_\omega.$$

Combining the latter relation with (23) leads to

$$v \in \mathrm{K}_{\hat{\sigma}} \quad \Rightarrow \quad \|A(\hat{u} + v) - d\|^2 \geqslant \|A\hat{u} - d\|^2 \ . \tag{24}$$

Let ρ be defined as in (19) Lemma 2.1. Noticing that by (2) and (17)

$$v \in \mathrm{K}_{\hat{\sigma}} \quad \Rightarrow \quad \phi\left(v[i]\right) = 0 \quad \forall\, i \in \hat{\sigma}^c \ , \tag{25}$$

the following inequality chain is derived:

$$\begin{aligned} v \in \mathrm{B}_\infty(0,\rho) \cap \mathrm{K}_{\hat{\sigma}} \quad \Rightarrow \quad \mathcal{F}_d(\hat{u} + v) &= \|A(\hat{u} + v) - d\|^2 + \beta \sum_{i \in \mathbb{I}_{\mathsf{N}}} \phi\left(\hat{u}[i] + v[i]\right) \\ \Big[\text{by Lemma 2.1(i)}\Big] \qquad &= \|A(\hat{u} + v) - d\|^2 + \beta \sum_{i \in \hat{\sigma}} \phi\left(\hat{u}[i]\right) + \beta \sum_{i \in \hat{\sigma}^c} \phi\left(v[i]\right) \\ \Big[\text{by (25)}\Big] \qquad &= \|A(\hat{u} + v) - d\|^2 + \beta \sum_{i \in \hat{\sigma}} \phi\left(\hat{u}[i]\right) \\ \Big[\text{by (24)}\Big] \qquad &\geqslant \|A\hat{u} - d\|^2 + \beta \sum_{i \in \hat{\sigma}} \phi\left(\hat{u}[i]\right) \\ \Big[\text{by (3)}\Big] \qquad &= \mathcal{F}_d(\hat{u}) \ . \end{aligned}$$

Combining the obtained implication with Lemma 2.1(ii) shows that

$$\mathcal{F}_d(\hat{u}+v) \geqslant \mathcal{F}_d(\hat{u}) \quad \forall\, v \in \mathrm{B}_\infty(0,\rho)\ .$$

If $\hat{u}=0$, this is a (local) minimizer of $\mathcal{F}_d$ by Lemma 2.2. □

Many authors mention that initialization is paramount for the success of approximate algorithms minimizing $\mathcal{F}_d$. In view of Proposition 2.3, if one already has a well-elaborated initialization, it could be enough to solve the relevant problem $(\mathcal{P}_\omega)$.

The statement reciprocal to Proposition 2.3 is obvious but it helps the presentation.

Lemma 2.4. *For $d \in \mathbb{R}^{\mathsf{M}}$ and $\beta > 0$, let $\mathcal{F}_d$ have a (local) minimum at $\hat{u}$. Then $\hat{u}$ solves $(\mathcal{P}_{\hat{\sigma}})$ for $\hat{\sigma} \stackrel{\text{def}}{=} \sigma(\hat{u})$.*

Proof. Let $\hat{u}$ be a (local) minimizer of $\mathcal{F}_d$. Denote $\hat{\sigma} \stackrel{\text{def}}{=} \sigma(\hat{u})$. Then $\hat{u}$ solves the problem

$$\min_{u\in\mathbb{R}^{\mathsf{N}}} \left\{ \|Au-d\|^2 + \beta\,\sharp\,\hat{\sigma} \right\} \ \text{ subject to } u[i]=0 \ \ \forall\, i \in \hat{\sigma}^c.$$

Since $\sharp\,\hat{\sigma}$ is a constant, $\hat{u}$ solves $(\mathcal{P}_{\hat{\sigma}})$. □

Remark 2. By Proposition 2.3 and Lemma 2.4, solving $(\mathcal{P}_\omega)$ for some $\omega \subseteq \mathbb{I}_{\mathsf{N}}$ is equivalent *to finding a (local) minimizer of $\mathcal{F}_d$.*

This equivalence underlies most of the theory developed in this work.

Corollary 2.5. *For $d \in \mathbb{R}^{\mathsf{M}}$ and $\beta > 0$, let $\hat{u}$ be a (local) minimizer of $\mathcal{F}_d$. Set $\hat{\sigma} \stackrel{\text{def}}{=} \sigma(\hat{u})$. Then*

$$\hat{u} = Z_{\hat{\sigma}}(\hat{u}_{\hat{\sigma}})\ , \quad \text{where } \ \hat{u}_{\hat{\sigma}} \ \text{ satisfies } \quad A_{\hat{\sigma}}^T A_{\hat{\sigma}} \hat{u}_{\hat{\sigma}} = A_{\hat{\sigma}}^T d\ . \tag{26}$$

Conversely, if $\hat{u} \in \mathbb{R}^{\mathsf{N}}$ satisfies (26) for $\hat{\sigma} = \sigma(\hat{u})$, then $\hat{u}$ is a (local) minimizer of $\mathcal{F}_d$.

Proof. By Lemma 2.4, $\hat{u}$ solves $(\mathcal{P}_{\hat{\sigma}})$. The equation in (26) follows directly from (16). The last claim is a straightforward consequence of (16) and Proposition 2.3. □

Remark 3. Equation (26) shows that a (local) minimizer $\hat{u}$ of $\mathcal{F}_d$ follows a *pseudo*-hard thresholding scheme[9]: the nonzero part $\hat{u}_{\hat{\sigma}}$ of $\hat{u}$ is the least squares solution with respect to the submatrix $A_{\hat{\sigma}}$ and the whole data vector d is involved in its computation. Unlike the hard thresholding scheme in [15], unsignificant or purely noisy data entries can hardly be discarded from $\hat{u}$ and they threaten to pollute its nonzero part $\hat{u}_{\hat{\sigma}}$. See also Remark 6.

Noisy data d should degrade $\hat{u}_{\hat{\sigma}}$ and this effect is stronger if $A_{\hat{\sigma}}^T A_{\hat{\sigma}}$ is ill-conditioned [13]. The quality of the outcome critically depends on the selected (local) minimizer and on the pertinence of A.

It may be interesting to evoke another consequence of Proposition 2.3:

Remark 4. Given $d \in \mathbb{R}^{\mathsf{M}}$, for any $\omega \subseteq \mathbb{I}_{\mathsf{N}}$, $\mathcal{F}_d$ has a (local) minimizer $\hat{u}$ defined by (26) and obeying $\sigma(\hat{u}) \subseteq \omega$.

[9] In a Bayesian setting, the quadratic data fidelity term in $\mathcal{F}_d$ models data corrupted with Gaussian i.i.d. noise.

3. The strict minimizers of $\mathcal{F}_d$.

We remind, yet again, that no special assumptions on $A \in \mathbb{R}^{\mathsf{M}\times\mathsf{N}}$ are adopted.

Strict minimizers of an objective function enable unambiguous solutions of inverse problems. The definition below is useful in characterizing the strict minimizers of $\mathcal{F}_d$.

Definition 3.1. *Given a matrix $A \in \mathbb{R}^{\mathsf{M}\times\mathsf{N}}$, for any $r \in \mathbb{I}_{\mathsf{M}}$ we define Ω_r as the subset of* all *r-length supports that correspond to full column rank $\mathsf{M} \times r$ submatrices of A, i.e.,*

$$\boxed{\Omega_r = \Big\{ \ \omega \subset \mathbb{I}_{\mathsf{N}} \ \ : \ \ \sharp\,\omega = r = \operatorname{rank}(A_\omega) \ \Big\} \ .}$$

Set $\Omega_0 = \varnothing$ and define as well

$$\boxed{\Omega \overset{\text{def}}{=} \bigcup_{r=0}^{\mathsf{M}-1} \Omega_r \quad \text{and} \quad \Omega_{\max} \overset{\text{def}}{=} \Omega \cup \Omega_{\mathsf{M}} \ .}$$

Definition 3.1 shows that for any $r \in \mathbb{I}_{\mathsf{M}}$,

$$\operatorname{rank}(A) = r \geqslant 1 \quad \Leftrightarrow \quad \Omega_r \neq \varnothing \ \text{ and } \ \Omega_t = \varnothing \ \ \forall\, t \geqslant r+1 \ .$$

3.1. How to recognize a strict minimizer of $\mathcal{F}_d$?

The theorem below gives an exhaustive answer to this question.

Theorem 3.2. *Given $d \in \mathbb{R}^{\mathsf{M}}$ and $\beta > 0$, let $\hat{u}$ be a (local) minimizer of $\mathcal{F}_d$. Define*

$$\hat{\sigma} \overset{\text{def}}{=} \sigma(\hat{u}) \ .$$

The following statements are equivalent*:*

(i) *The (local) minimum that $\mathcal{F}_d$ has at $\hat{u}$ is* strict*;*

(ii) $\boxed{\operatorname{rank}(A_{\hat{\sigma}}) = \sharp\,\hat{\sigma}}$ *;*

(iii) $\hat{\sigma} \in \Omega_{\max}$.

If $\hat{u}$ is a strict (local) minimizer of $\mathcal{F}_d$, then it reads

$$\hat{u} = Z_{\hat{\sigma}}(\hat{u}_{\hat{\sigma}}) \quad \textit{for} \quad \hat{u}_{\hat{\sigma}} = \left(A_{\hat{\sigma}}^T A_{\hat{\sigma}}\right)^{-1} A_{\hat{\sigma}}^T d \tag{27}$$

and satisfies $\sharp\,\hat{\sigma} = \|\hat{u}\|_0 \leqslant \mathsf{M}$.

Proof. We break the proof into four parts.

[(i)$\Rightarrow$(ii)]. We recall that by the rank-nullity theorem [21, 29]

$$\dim\ker(A_{\hat{\sigma}}) = \sharp\,\hat{\sigma} - \operatorname{rank}(A_{\hat{\sigma}}) \ . \tag{28}$$

Let[10] $\hat{u} \neq 0$ be a *strict* (local) minimizer of $\mathcal{F}_d$. Assume that (ii) fails. Then (28) implies that

$$\dim\ker(A_{\hat{\sigma}}) \geqslant 1 \ . \tag{29}$$

[10]This part can alternatively be proven using Remark 1.

By Lemma 2.4, $\hat{u}$ solves $(\mathcal{P}_{\hat{\sigma}})$. Let ρ read as in (19) and let $\mathrm{K}_{\hat{\sigma}}$ be defined according to (17). Noticing that

$$v \in \mathrm{K}_{\hat{\sigma}}, \quad \hat{\sigma} \neq \varnothing \quad \Rightarrow \quad Av = A_{\hat{\sigma}} v_{\hat{\sigma}} \ , \tag{30}$$

Lemma 2.1(i) shows that

$$\begin{aligned} \left\{ \begin{array}{l} v \in \mathrm{B}_\infty(0,\rho) \cap \mathrm{K}_{\hat{\sigma}} \ , \\ \\ \quad v_{\hat{\sigma}} \in \ker(A_{\hat{\sigma}}) \end{array} \right. \quad \Rightarrow \quad \mathcal{F}_d(\hat{u}+v) &= \|A_{\hat{\sigma}}(\hat{u}_{\hat{\sigma}} + v_{\hat{\sigma}}) - d\|^2 + \beta \sum_{i\in\hat{\sigma}} \phi(\hat{u}[i] + v[i]) \\ \Big[\text{by Lemma 2.1(i)}\Big] \qquad &= \|A_{\hat{\sigma}}\hat{u}_{\hat{\sigma}} - d\|^2 + \beta \sum_{i\in\hat{\sigma}} \phi(\hat{u}[i]) + \beta \sum_{i\in\hat{\sigma}^c} \phi(v[i]) \\ \Big[\text{by (25)}\Big] \qquad &= \|A_{\hat{\sigma}}\hat{u}_{\hat{\sigma}} - d\|^2 + \beta \sum_{i\in\hat{\sigma}} \phi(\hat{u}[i]) \\ \Big[\text{by (3)}\Big] \qquad &= \mathcal{F}_d(\hat{u}) \ , \end{aligned}$$

i.e., that $\hat{u}$ is not a strict minimizer, which contradicts (i). Hence the assumption in (29) is false. Therefore (ii) holds true.

If $\hat{u} = 0$, then $\hat{\sigma} = \varnothing$; hence $A_{\hat{\sigma}} \in \mathbb{R}^{\mathsf{M}\times 0}$ and $\mathrm{rank}\,(A_{\hat{\sigma}}) = 0 = \sharp\,\hat{\sigma}$ according to (13).

[(ii) ⇒ (i)]. Let $\hat{u}$ be a minimizer of $\mathcal{F}_d$ that satisfies (ii). To have $\sharp\,\hat{\sigma} = 0$ is equivalent to $\hat{u} = 0$. By Lemma 2.2, $\hat{u}$ is a strict minimizer. Focus on $\sharp\,\hat{\sigma} \geqslant 1$. Since $\mathrm{rank}\,(A_{\hat{\sigma}}) = \sharp\,\hat{\sigma} \leqslant \mathsf{M}$ and problem $(\mathcal{Q}_\omega)$ in (15) is strictly convex for $\omega = \hat{\sigma}$, its unique solution $\hat{u}_{\hat{\sigma}}$ satisfies

$$v \in \mathbb{R}^{\sharp\,\hat{\sigma}} \setminus \{0\} \quad \Rightarrow \quad \|A_{\hat{\sigma}}(\hat{u}_{\hat{\sigma}} + v) - d\|^2 > \|A_{\hat{\sigma}}\hat{u}_{\hat{\sigma}} - d\|^2 \ .$$

Using (30), this is equivalent to

$$v \in \mathrm{K}_{\hat{\sigma}} \setminus \{0\} \quad \Rightarrow \quad \|A(\hat{u}+v) - d\|^2 = \|A_{\hat{\sigma}}(\hat{u}_{\hat{\sigma}} + v_{\hat{\sigma}}) - d\|^2 > \|A_{\hat{\sigma}}\hat{u}_{\hat{\sigma}} - d\|^2 = \|A\hat{u} - d\|^2 \ . \tag{31}$$

Lemma 2.1(i), along with (25), yields

$$\begin{aligned} v \in \mathrm{B}_\infty(0,\rho) \cap \mathrm{K}_{\hat{\sigma}} \setminus \{0\} \quad \Rightarrow \quad \mathcal{F}_d(\hat{u}+v) &= \|A(u+v) - d\|^2 + \beta \sum_{i\in\hat{\sigma}} \phi(\hat{u}[i]) \\ \Big[\text{by (31)}\Big] \qquad &> \|A\hat{u} - d\|^2 + \beta \sum_{i\in\hat{\sigma}} \phi(\hat{u}[i]) \\ &= \mathcal{F}_d(\hat{u}) \ . \end{aligned}$$

Since $\sharp\,\hat{\sigma} \leqslant \mathsf{M} \leqslant \mathsf{N} - 1$, we have $\hat{\sigma}^c \neq \varnothing$. So Lemma 2.1(ii) tells us that

$$v \in \mathrm{B}_\infty(0,\rho) \setminus \mathrm{K}_{\hat{\sigma}} \quad \Rightarrow \quad \mathcal{F}_d(\hat{u}+v) > \mathcal{F}_d(\hat{u}) \ .$$

Combining the last two implications proves (i).

[(ii) ⇒ (iii)]. Comparing (iii) with Definitions 1.1 and 3.1 proves the claim.

[Equation (27)]. The proof follows from equation (26) in Corollary 2.5 where[11] $A_{\hat{\sigma}}^T A_{\hat{\sigma}}$ is invertible. $\square$

Theorem 3.2 provides a simple rule enabling one to verify whether or not a numerical scheme has reached a strict (local) minimizer of $\mathcal{F}_d$.

The notations Ω_r, Ω and $\Omega_{\max}$ are frequently used in this paper. Their interpretation is obvious in the light of Theorem 3.2. For any $d \in \mathbb{R}^{\mathsf{M}}$ and for all $\beta > 0$, the set $\Omega_{\max}$ is composed of the supports of all possible strict (local) minimizers of $\mathcal{F}_d$, while Ω is is the subset of those that are $(\mathsf{M}-1)$-sparse.

An easy and useful corollary is presented next.

Corollary 3.3. *Let $d \in \mathbb{R}^{\mathsf{M}}$. Given an arbitrary $\omega \in \Omega_{\max}$, let $\hat{u}$ solve $(\,\mathcal{P}_\omega\,)$. Then*

(i) *$\hat{u}$ reads as*

$$\hat{u} = Z_\omega\left(\hat{u}_\omega\right) \ , \quad \text{where} \quad \hat{u}_\omega = \left(A_\omega^T A_\omega\right)^{-1} A_\omega^T d \ , \tag{32}$$

and obeys $\hat{\sigma} \stackrel{\text{def}}{=} \sigma(\hat{u}) \subseteq \omega$ and $\hat{\sigma} \in \Omega_{\max}$;

(ii) *for any $\beta > 0$, $\hat{u}$ is a* strict *(local) minimizer of $\mathcal{F}_d$;*

(iii) *$\hat{u}$ solves $(\,\mathcal{P}_{\hat{\sigma}}\,)$.*

Proof. Using (16), $\hat{u}$ fulfills (i) since $A_\omega^T A_\omega$ is invertible and $\sigma(\hat{u}) \subseteq \omega$ by the constraint in $(\,\mathcal{P}_\omega\,)$. If $\hat{\sigma} = \varnothing$, (ii) follows from Lemma 2.2. For $\sharp\,\hat{\sigma} \geqslant 1$, $A_{\hat{\sigma}}$ is an $\mathsf{M} \times \sharp\,\hat{\sigma}$ submatrix of A_ω. Since $\operatorname{rank}(A_\omega) = \sharp\,\omega$, we have $\operatorname{rank}(A_{\hat{\sigma}}) = \sharp\,\hat{\sigma}$ and so $\hat{\sigma} \in \Omega_{\max}$. By Proposition 2.3 $\hat{u}$ is a (local) minimizer of $\mathcal{F}_d$, and Theorem 3.2 leads to (ii). Lemma 2.4 and Corollary 3.3(ii) yield (iii). $\square$

Remark 5. One can easily compute a strict (local) minimizer $\hat{u}$ of $\mathcal{F}_d$ without knowing the value of the regularization parameter β. Just data d and an $\omega \in \Omega_{\max}$ are needed.

This consequence of Corollary 3.3 might be striking.

Clearly, the support $\sigma(\bar{u})$ of a nonstrict local minimizer $\bar{u}$ of $\mathcal{F}_d$ contains some subsupports yielding strict (local) minimizers of $\mathcal{F}_d$. It is easy to see that among them, there is $\hat{\sigma} \subsetneqq \sigma(\bar{u})$ such that the corresponding $\hat{u}$ given by (27) *strictly* decreases the value of $\mathcal{F}_d$; i.e., $\mathcal{F}_d(\hat{u}) < \mathcal{F}_d(\bar{u})$.

3.2. Every strict (local) minimizer of $\mathcal{F}_d$ is linear in d.

Here we explore the behavior of the strict (local) minimizers of $\mathcal{F}_d$ with respect to variations of d. An interesting sequel of Theorem 3.2 is presented in the following corollary.

Corollary 3.4. *For $d \in \mathbb{R}^{\mathsf{M}}$ and $\beta > 0$, let $\hat{u}$ be a (local) minimizer of $\mathcal{F}_d$ satisfying $\hat{\sigma} \stackrel{\text{def}}{=} \sigma(\hat{u}) \in \Omega$. Define*

$$\mathrm{N}_{\hat{\sigma}} \stackrel{\text{def}}{=} \ker\left(A_{\hat{\sigma}}^T\right) \subset \mathbb{R}^{\mathsf{M}} \ .$$

We have $\dim \mathrm{N}_{\hat{\sigma}} = \mathsf{M} - \sharp\,\hat{\sigma} \geqslant 1$ and

$$d' \in \mathrm{N}_{\hat{\sigma}} \quad \Rightarrow \quad \mathcal{F}_{d+d'} \ \text{ has a strict (local) minimum at } \ \hat{u} \ .$$

[11] For $\hat{\sigma} = \varnothing$, (11) and (13)$(a)$ show that (27) yields $\hat{u} = 0$.

Proof. Since $\hat\sigma \in \Omega$, the minimizer $\hat u$ is strict by Theorem 3.2. By $d' \in \ker\left(A_{\hat\sigma}^T\right)$ we find $A_{\hat\sigma}^T(d+d') = A_{\hat\sigma}^T d$ for any $d' \in \mathrm{N}_{\hat\sigma}$. Inserting this into (27) in Theorem 3.2 yields the result. ☐

All data located in the vector subspace $\mathrm{N}_{\hat\sigma} \supsetneqq \{0\}$ yield the same strict (local) minimizer $\hat u$.

Remark 6. If data contain noise n, it can be decomposed in a unique way as $n = n_{\mathrm{N}_{\hat\sigma}} + n_{\mathrm{N}_{\hat\sigma}^\perp}$ where $n_{\mathrm{N}_{\hat\sigma}} \in \mathrm{N}_{\hat\sigma}$ and $n_{\mathrm{N}_{\hat\sigma}^\perp} \in \mathrm{N}_{\hat\sigma}^\perp$. The component $n_{\mathrm{N}_{\hat\sigma}}$ is removed (Corollary 3.4), while $n_{\mathrm{N}_{\hat\sigma}^\perp}$ is transformed according to (27) and added to $\hat u_{\hat\sigma}$.

We shall use the following definition.

Definition 3.5. *Let $\mathcal{O} \subseteq \mathbb{R}^{\mathsf{M}}$ be an open domain. We say that $\mathcal{U} : \mathcal{O} \to \mathbb{R}^{\mathsf{N}}$ is a* local minimizer function *for the family of objectives $\mathfrak{F}_{\mathcal{O}} \stackrel{\text{def}}{=} \{\mathcal{F}_d \ : \ d \in \mathcal{O}\}$ if for any $d \in \mathcal{O}$, the function $\mathcal{F}_d$ reaches a* strict *(local) minimum at $\mathcal{U}(d)$.*

Corollary 3.3 shows that for any $d \in \mathbb{R}^{\mathsf{M}}$, each strict (local) minimizer of $\mathcal{F}_d$ is entirely described by an $\omega \in \Omega_{\max}$ via equation (32) in the same corollary. Consequently, a local minimizer function $\mathcal{U}$ is associated with every $\omega \in \Omega_{\max}$.

Lemma 3.6. *For some arbitrarily fixed $\omega \in \Omega_{\max}$ and $\beta > 0$, the family of functions $\mathfrak{F}_{\mathbb{R}^{\mathsf{M}}}$ has a* linear *(local) minimizer function $\mathcal{U} : \mathbb{R}^{\mathsf{M}} \to \mathbb{R}^{\mathsf{N}}$ that reads as*

$$\forall\, d \in \mathbb{R}^{\mathsf{M}}, \quad \mathcal{U}(d) = Z_\omega\left(U_\omega\, d\right) \ , \quad \textit{where} \ \ U_\omega = \left(A_\omega^T A_\omega\right)^{-1} A_\omega^T \ \in \mathbb{R}^{\sharp\omega \times \mathsf{M}} \ . \tag{33}$$

Proof. The function $\mathcal{U}$ in (33) is linear in d. From Corollary 3.3, for any $\beta > 0$ and for any $d \in \mathbb{R}^{\mathsf{M}}$, $\mathcal{F}_d$ has a strict (local) minimum at $\hat u = \mathcal{U}(d)$. Hence $\mathcal{U}$ fits Definition 3.5. ☐

Thus, even if $\mathcal{F}_d$ has many strict (local) minimizers, each is linear in d.

Next we exhibit a *closed negligible* subset of $\mathbb{R}^{\mathsf{M}}$, associated with a nonempty $\omega \in \Omega_{\max}$, whose elements are data d leading to $\|\mathcal{U}(d)\|_0 < \sharp\,\omega$.

Lemma 3.7. *For any $\omega \in \Omega_{\max}$, define the subset $\mathrm{D}_\omega \subset \mathbb{R}^{\mathsf{M}}$ by*

$$\mathrm{D}_\omega \stackrel{\text{def}}{=} \bigcup_{i=1}^{\sharp\,\omega} \left\{ \, g \in \mathbb{R}^{\mathsf{M}} \ : \ \left\langle e_i\,,\, \left(A_\omega^T A_\omega\right)^{-1} A_\omega^T\, g \right\rangle = 0 \, \right\} . \tag{34}$$

Then D_ω is closed in $\mathbb{R}^{\mathsf{M}}$ and $\mathbb{L}^{\mathsf{M}}\left(\mathrm{D}_\omega\right) = 0$.

Proof. If $\omega = \varnothing$ then $\mathrm{D}_\omega = \varnothing$, hence the claim. Let $\sharp\,\omega \geqslant 1$. For some $i \in \mathbb{I}_{\sharp\omega}$, set

$$\begin{aligned} \mathrm{D} &\stackrel{\text{def}}{=} \left\{ \, g \in \mathbb{R}^{\mathsf{M}} \ : \ \left\langle e_i\,,\, \left(A_\omega^T A_\omega\right)^{-1} A_\omega^T\, g \right\rangle = 0 \, \right\} \\ &= \left\{ \, g \in \mathbb{R}^{\mathsf{M}} \ : \ \left\langle A_\omega \left(A_\omega^T A_\omega\right)^{-1} e_i,\ g \right\rangle = 0 \right\} . \end{aligned}$$

Since $\operatorname{rank}\left(A_\omega \left(A_\omega^T A_\omega\right)^{-1}\right) = \sharp\,\omega$, $\ker\left(A_\omega \left(A_\omega^T A_\omega\right)^{-1}\right) = \{0\}$. Hence $A_\omega \left(A_\omega^T A_\omega\right)^{-1} e_i \neq 0$. Therefore D is a vector subspace of $\mathbb{R}^{\mathsf{M}}$ of dimension $\mathsf{M}-1$ and so $\mathbb{L}^{\mathsf{M}}(\mathrm{D}) = 0$. The conclusion follows from the fact that D_ω in (34) is the union of $\sharp\,\omega$ subsets like D (see, e.g., [36, 16]). ☐

Proposition 3.8. *For some arbitrarily fixed $\omega \in \Omega_{\max}$ and $\beta > 0$, let $\mathcal{U} : \mathbb{R}^{\mathsf{M}} \to \mathbb{R}^{\mathsf{N}}$ be the relevant (local) minimizer function for $\mathfrak{F}_{\mathbb{R}^{\mathsf{M}}}$ as given in* (33) *(Lemma 3.6). Let D_ω read as in* (34). *Then the function $d \mapsto \mathcal{F}_d\big(\mathcal{U}(d)\big)$ is $\mathcal{C}^\infty$ on $\mathbb{R}^{\mathsf{M}} \setminus \mathrm{D}_\omega$ and*

$$d \in \mathbb{R}^{\mathsf{M}} \setminus \mathrm{D}_\omega \quad \Rightarrow \quad \sigma\left(\mathcal{U}(d)\right) = \omega \ ,$$

where the set $\mathbb{R}^{\mathsf{M}} \setminus \mathrm{D}_\omega$ contains an open and dense subset of $\mathbb{R}^{\mathsf{M}}$.

Proof. The statement about $\mathbb{R}^{\mathsf{M}} \setminus \mathrm{D}_\omega$ is a direct consequence of Lemma 3.7.

If $\omega = \varnothing$, then $\mathcal{U}(d) = 0$ for all $d \in \mathbb{R}^{\mathsf{M}}$, so all claims in the proposition are trivial. Consider that $\sharp\,\omega \geqslant 1$. For any $i \in \mathbb{I}_{\sharp\,\omega}$, the $\omega[i]$th component of $\mathcal{U}(d)$ reads as (see Lemma 3.6)

$$\mathcal{U}_{\omega[i]}(d) = \left\langle e_i \,,\, \left(A_\omega^T A_\omega\right)^{-1} A_\omega^T\, d \right\rangle \ .$$

The definition of D_ω shows that

$$d \in \mathbb{R}^{\mathsf{M}} \setminus \mathrm{D}_\omega \ \text{ and } \ i \in \mathbb{I}_{\sharp\,\omega} \quad \Rightarrow \quad \mathcal{U}_{\omega[i]}(d) \neq 0 \ ,$$

whereas $\mathcal{U}_i(d) = 0$ for all $i \in \omega^c$. Consequently,

$$\omega \in \Omega_{\max} \ \text{ and } \ d \in \mathbb{R}^{\mathsf{M}} \setminus \mathrm{D}_\omega \quad \Rightarrow \quad \sigma\left(\mathcal{U}(d)\right) = \omega \ .$$

Then $\|\mathcal{U}(d)\|_0$ is constant on $\mathbb{R}^{\mathsf{M}} \setminus \mathrm{D}_\omega$ and

$$\omega \in \Omega_{\max} \ \text{ and } \ d \in \mathbb{R}^{\mathsf{M}} \setminus \mathrm{D}_\omega \quad \Rightarrow \quad \mathcal{F}_d(\mathcal{U}(d)) = \big\|A\mathcal{U}(d) - d\big\|^2 + \beta\,\sharp\,\omega \ .$$

We infer from (33) that $d \mapsto \big\|A\mathcal{U}(d) - d\big\|^2$ is $\mathcal{C}^\infty$ on $\mathbb{R}^{\mathsf{M}}$, so $d \mapsto \mathcal{F}_d\big(\mathcal{U}(d)\big)$ is $\mathcal{C}^\infty$ on $\mathbb{R}^{\mathsf{M}} \setminus \mathrm{D}_\omega$. □

A generic property is that a local minimizer function corresponding to $\mathcal{F}_d$ produces solutions sharing the same support. The application $d \mapsto \mathcal{F}_d\big(\mathcal{U}(d)\big)$ is discontinuous on the *closed negligible* subset D_ω, where the support of $\mathcal{U}(d)$ is shrunk.

3.3. Strict minimizers with an M-length support. Here we explain why minimizers with an M-length support are useless in general.

Proposition 3.9. *Let* $\operatorname{rank}(A) = \mathsf{M}$, $\beta > 0$ *and for* $d \in \mathbb{R}^{\mathsf{M}}$ *set*

$$\mathrm{U}_{\mathsf{M}} \stackrel{\text{def}}{=} \left\{ \hat{u} \in \mathbb{R}^{\mathsf{N}} \ : \ \hat{u} \ \ \textit{is a strict (local) minimizer of} \ \ \mathcal{F}_d \ \ \textit{meeting} \ \ \|\hat{u}\|_0 = \mathsf{M} \ \right\} \ .$$

Put

$$\mathrm{Q}_{\mathsf{M}} \stackrel{\text{def}}{=} \bigcup_{\omega \in \Omega_{\mathsf{M}}} \ \bigcup_{i \in \mathbb{I}_{\mathsf{M}}} \left\{ g \in \mathbb{R}^{\mathsf{M}} \ : \ \left\langle e_i \,,\, A_\omega^{-1} g \right\rangle = 0 \right\} \ . \tag{35}$$

Then $\mathbb{R}^{\mathsf{M}} \setminus \mathrm{Q}_{\mathsf{M}}$ contains a dense open subset of $\mathbb{R}^{\mathsf{M}}$ and

$$d \in \mathbb{R}^{\mathsf{M}} \setminus \mathrm{Q}_{\mathsf{M}} \quad \Rightarrow \quad \sharp\, \mathrm{U}_{\mathsf{M}} = \sharp\, \Omega_{\mathsf{M}} \quad \text{and} \quad \mathcal{F}_d(\hat{u}) = \beta \mathsf{M} \quad \forall\, \hat{u} \in \mathrm{U}_{\mathsf{M}} \ .$$

Proof. Using the notation in (34), Q_{M} reads as

$$\mathrm{Q}_{\mathsf{M}} = \bigcup_{\omega\in\Omega_{\mathsf{M}}} \mathrm{D}_\omega \ .$$

The claim on $\mathbb{R}^{\mathsf{M}} \setminus \mathrm{Q}_{\mathsf{M}}$ follows from Lemma 3.7. Since $\mathrm{rank}(A) = \mathsf{M}$, we have $\sharp\,\Omega_{\mathsf{M}} \geqslant 1$.

Consider that $d \in \mathbb{R}^{\mathsf{M}} \setminus \mathrm{Q}_{\mathsf{M}}$. By Proposition 3.8

$$d \in \mathbb{R}^{\mathsf{M}} \setminus \mathrm{Q}_{\mathsf{M}} \quad \text{and} \quad \omega \in \Omega_{\mathsf{M}} \quad \Rightarrow \quad \mathcal{F}_d \text{ has a strict (local) minimizer } \ \hat{u} \ \text{ obeying } \sigma(\hat{u}) = \omega \ .$$

Hence $\hat{u} \in \mathrm{U}_{\mathsf{M}}$. Therefore, we have a mapping $b : \Omega_{\mathsf{M}} \to \mathrm{U}_{\mathsf{M}}$ such that $\hat{u} = b(\omega) \in \mathrm{U}_{\mathsf{M}}$. Using Lemma 2.4 and Corollary 3.3, it reads as

$$b(\omega) = Z_\omega(A_\omega^{-1} d) \ .$$

For $(\omega, \varpi) \in \Omega_{\mathsf{M}} \times \Omega_{\mathsf{M}}$ with $\varpi \neq \omega$ one obtains $\hat{u} = b(\omega) \in \mathrm{U}_{\mathsf{M}}$, $\bar{u} = b(\varpi) \in \mathrm{U}_{\mathsf{M}}$ and $\hat{u} \neq \bar{u}$, hence b is one-to-one. Conversely, for any $\hat{u} \in \mathrm{U}_{\mathsf{M}}$ there is $\omega \in \Omega_{\mathsf{M}}$ such that $\hat{u} = b(\omega)$ and $\sigma(\hat{u}) = \omega$ (because $d \notin \mathrm{Q}_{\mathsf{M}}$). It follows that b maps Ω_{M} onto U_{M}. Therefore, Ω_{M} are U_{M} in one-to-one correspondence, i.e. $\sharp\,\Omega_{\mathsf{M}} = \sharp\,\mathrm{U}_{\mathsf{M}}$.

Last, it is clear that $\omega \in \Omega_{\mathsf{M}}$ and $d \notin \mathrm{Q}_{\mathsf{M}}$ lead to $\|A\hat{u} - d\|^2 = 0$ and $\mathcal{F}_d(\hat{u}) = \beta\mathsf{M}$. □

For any $\beta > 0$, a generic property of $\mathcal{F}_d$ is that it has $\sharp\,\Omega_{\mathsf{M}}$ strict minimizers $\hat{u}$ obeying $\|\hat{u}\|_0 = \mathsf{M}$ and $\mathcal{F}_d(\hat{u}) = \beta\mathsf{M}$. It is hard to discriminate between all these minimizers. Hence the interest in minimizers with supports located in Ω, i.e., strict $(\mathsf{M}-1)$-sparse minimizers of $\mathcal{F}_d$.

4. On the global minimizers of $\mathcal{F}_d$. The next proposition gives a *necessary condition* for a global minimizer of $\mathcal{F}_d$. It follows directly from [32, Proposition 3.4] where[12] the regularization term is $\|Du\|_0$ for a full row rank matrix D. For $\mathcal{F}_d$ in (1) with $\|a_i\|_2 = 1$, $\forall\, i \in \mathbb{I}_{\mathsf{N}}$, a simpler condition was derived later in [40, Theorem 12], using different tools. For completeness, the proof for a general A is outlined in Appendix 8.2.

Proposition 4.1. *For $d \in \mathbb{R}^{\mathsf{M}}$ and $\beta > 0$, let $\mathcal{F}_d$ have a global minimum at $\hat{u}$. Then*[13]

$$i \in \sigma(\hat{u}) \quad \Rightarrow \quad \big|\,\hat{u}[i]\,\big| \geqslant \frac{\sqrt{\beta}}{\|a_i\|} \ . \tag{36}$$

Observe that the lower bound on $\big\{\big|\,\hat{u}[i]\,\big| \ : \ i \in \sigma(\hat{u})\big\}$ given in (36) is independent of d. This means that in general, (36) provides a pessimistic bound.

The proof of the statement shows that (36) is met also by all (local) minimizers of $\mathcal{F}_d$ satisfying

$$\mathcal{F}_d(\hat{u}) \leqslant \mathcal{F}_d(\hat{u} + \rho e_i) \quad \forall\, \rho \in \mathbb{R}, \quad \forall\, i \in \mathbb{I}_{\mathsf{N}} \ .$$

[12] Just set $g_i = e_i$, $P = I_{\mathsf{M}}$ and $H = I_{\mathsf{N}}$ in [32, Proposition 3.4].

[13] Recall that $a_i \neq 0$ for all $i \in \mathbb{I}_{\mathsf{N}}$ by (8) and that $\|\cdot\| = \|\cdot\|_2$.

4.1. The global minimizers of $\mathcal{F}_d$ are strict.

Remark 7. Let $d \in \mathbb{R}^{\mathsf{M}}$ and $\beta > \|d\|^2$. Then $\mathcal{F}_d$ has a strict global minimum at $\hat{u} = 0$. Indeed,

$$v \neq 0 \quad \Rightarrow \quad \|v\|_0 \geqslant 1 \quad \Rightarrow \quad \mathcal{F}_d(0) = \|d\|^2 < \beta \leqslant \|Av - d\|^2 + \beta\|v\|_0 \ .$$

For least-squares regularized with a more regular ϕ, one usually gets $\hat{u} = 0$ asymptotically as $\beta \to +\infty$ but $\hat{u} \neq 0$ for finite values of β. This does not hold for $\mathcal{F}_d$ by Remark 7.

Some theoretical results on the global minimizers of $\mathcal{F}_d$ have been obtained [32, 22, 40, 7]. Surprisingly, the question about the *existence of global minimizers of* $\mathcal{F}_d$ has never been raised. We answer this question using the notion of *asymptotically level stable functions* introduced by Auslender [2] in 2000. As usual,

$$\text{lev}\,(\mathcal{F}_d, \lambda) \stackrel{\text{def}}{=} \{v \in \mathbb{R}^{\mathsf{N}} \ : \ \mathcal{F}_d(v) \leqslant \lambda\} \quad \text{for} \quad \lambda > \inf \mathcal{F}_d \ .$$

The following definition is taken from [3, p. 94].

Definition 4.2. *Let $\mathcal{F}_d : \mathbb{R}^{\mathsf{N}} \to \mathbb{R} \cup \{+\infty\}$ be lower semicontinuous and proper. Then $\mathcal{F}_d$ is said to be* asymptotically level stable *if for each $\rho > 0$, each bounded sequence $\{\lambda_k\} \in \mathbb{R}$ and each sequence $\{v_k\} \in \mathbb{R}^{\mathsf{N}}$ satisfying*

$$v_k \in \text{lev}\,(\mathcal{F}_d, \lambda_k), \quad \|v_k\| \to +\infty, \quad v_k\,\|v_k\|^{-1} \to \bar{v} \in \ker\left((\mathcal{F}_d)_\infty\right) \ , \tag{37}$$

where $(\mathcal{F}_d)_\infty$ denotes the asymptotic (or recession) function of $\mathcal{F}_d$, there exists k_0 such that

$$v_k - \rho\bar{v} \in \text{lev}\,(\mathcal{F}_d, \lambda_k) \quad \forall\, k \geqslant k_0 \ . \tag{38}$$

One can note that a coercive function is asymptotically level stable, since (37) is empty. We prove that our discontinuous noncoercive objective $\mathcal{F}_d$ is asymptotically level stable as well.

Proposition 4.3. *Let $\mathcal{F}_d : \mathbb{R}^{\mathsf{N}} \to \mathbb{R}$ be of the form (1). Then $\ker\left((\mathcal{F}_d)_\infty\right) = \ker(A)$ and $\mathcal{F}_d$ is asymptotically level stable.*

The proof is outlined in Appendix 8.3.

Theorem 4.4.*Let $d \in \mathbb{R}^{\mathsf{M}}$ and $\beta > 0$. Then*

(i) *the set of the global minimizers of $\mathcal{F}_d$*

$$\hat{U} \stackrel{\text{def}}{=} \left\{\hat{u} \in \mathbb{R}^{\mathsf{N}} \ : \ \hat{u} = \min_{u \in \mathbb{R}^{\mathsf{N}}} \mathcal{F}_d(u)\right\} \tag{39}$$

is nonempty;

(ii) *every $\hat{u} \in \hat{U}$ is a* strict *minimizer of $\mathcal{F}_d$, i.e.,*

$$\sigma(\hat{u}) \in \Omega_{\max} \ ,$$

hence $\|\hat{u}\|_0 \leqslant \mathsf{M}$.

Proof. For claim (i), we use the following statement[14], whose proof can be found in the monograph by Ausleneder and Teboulle [3]:
[3, Corollary 3.4.2] *Let* $\mathcal{F}_d : \mathbb{R}^{\mathsf{N}} \to \mathbb{R} \cup \{+\infty\}$ *be asymptotically level stable with* $\inf \mathcal{F}_d > -\infty$. *Then the optimal set* $\hat{U}$ *—as given in* (39)*—is nonempty .*
From Proposition 4.3, $\mathcal{F}_d$ is asymptotically level stable and $\inf \mathcal{F}_d \geqslant 0$ from (1). Hence $\hat{U} \neq \varnothing$.

(ii). Let $\hat{u}$ be a global minimizer of $\mathcal{F}_d$. Set $\hat{\sigma} = \sigma(\hat{u})$.

If $\hat{u} = 0$, (ii) follows from Lemma 2.2. Suppose that the global minimizer $\hat{u} \neq 0$ of $\mathcal{F}_d$ is *nonstrict*. Then Theorem 3.2(ii) fails to hold; i.e.,

$$\dim \ker (A_{\hat{\sigma}}) \geqslant 1 \ . \tag{40}$$

Choose $v_{\hat{\sigma}} \in \ker (A_{\hat{\sigma}}) \setminus \{0\}$ and set $v = Z_{\hat{\sigma}}(v_{\hat{\sigma}})$. Select an $i \in \hat{\sigma}$ obeying $v[i] \neq 0$. Define $\widetilde{u}$ by

$$\widetilde{u} \stackrel{\text{def}}{=} \hat{u} - \hat{u}[i] \frac{v}{v[i]} \ . \tag{41}$$

We have $\widetilde{u}[i] = 0$ and $\hat{u}[i] \neq 0$. Set $\widetilde{\sigma} \stackrel{\text{def}}{=} \sigma(\widetilde{u})$. Then

$$\widetilde{\sigma} \subsetneqq \hat{\sigma} \quad \text{hence} \quad \sharp\,\widetilde{\sigma} \leqslant \sharp\,\hat{\sigma} - 1 \ . \tag{42}$$

From $v_{\hat{\sigma}} \dfrac{\hat{u}[i]}{v[i]} \in \ker (A_{\hat{\sigma}})$, using (12) and Remark 1 shows that[15] $A\hat{u} = A_{\hat{\sigma}}\hat{u}_{\hat{\sigma}} = A_{\widetilde{\sigma}}\widetilde{u}_{\widetilde{\sigma}} = A\widetilde{u}$. Then

$$\begin{aligned}\mathcal{F}_d(\widetilde{u}) &= \|A\widetilde{u} - d\|^2 + \beta\,\sharp\,\widetilde{\sigma} \\ &\leqslant \mathcal{F}_d(\hat{u}) - \beta = \|A\hat{u} - d\|^2 + \beta\,(\sharp\,\hat{\sigma} - 1) \ .\end{aligned}$$

It follows that $\hat{u}$ is not a global minimizer, hence (40) is false. Therefore $\operatorname{rank}(A_{\hat{\sigma}}) = \sharp\,\hat{\sigma}$ and $\hat{u}$ is a strict minimizer of $\mathcal{F}_d$ (Theorem 3.2). □

One can note that if $\operatorname{rank}(A) = \mathsf{M}$, any global minimizer $\hat{u}$ of $\mathcal{F}_d$ obeys $\mathcal{F}_d(\hat{u}) \leqslant \beta\mathsf{M}$.

According to Theorem 4.4, *the global minimizers of* $\mathcal{F}_d$ *are strict and their number is finite: this is a nice property that fails for many convex nonsmooth optimization problems.*

4.2. K-sparse global minimizers for $\mathsf{K} \leqslant \mathsf{M} - 1$.

In order to simplify the presentation, in what follows we consider that

$$\boxed{\operatorname{rank}(A) = \mathsf{M} < \mathsf{N} \ .}$$

Since $\mathcal{F}_d$ has a large number (typically equal to $\sharp\,\Omega_{\mathsf{M}}$) of strict minimizers with $\|\hat{u}\|_0 = \mathsf{M}$ yielding the same value $\mathcal{F}_d(\hat{u}) = \beta\mathsf{M}$ (see Proposition 3.9 and the comments given after its proof), it is important to be sure that the global minimizers of $\mathcal{F}_d$ are $(\mathsf{M}-1)$-sparse.

[14] This result was originally exhibited in [4] (without the notion of asymptotically level stable functions).

[15] In detail we have $A\hat{u} = A_{\hat{\sigma}}\hat{u}_{\hat{\sigma}} = A_{\hat{\sigma}}\left(\hat{u}_{\hat{\sigma}} - v_{\hat{\sigma}}\frac{\hat{u}[i]}{v[i]}\right) = A_{\hat{\sigma}}\widetilde{u}_{\hat{\sigma}} = A_{\widetilde{\sigma}}\widetilde{u}_{\widetilde{\sigma}} = A\widetilde{u}$.

We introduce a notation which is used in the rest of this paper. For any $\mathsf{K} \in \mathbb{I}_{\mathsf{M}-1}$, put

$$\boxed{\overline{\Omega}_{\mathsf{K}} \stackrel{\text{def}}{=} \bigcup_{r=0}^{\mathsf{K}} \Omega_r} \tag{43}$$

where Ω_r was set up in Definition 3.1. Theorem 3.2 gives a clear meaning of the sets[16] $\overline{\Omega}_{\mathsf{K}}$. *For any $d \in \mathbb{R}^{\mathsf{M}}$ and any $\beta > 0$, for any fixed $\mathsf{K} \in \mathbb{I}_{\mathsf{M}-1}$, the set $\overline{\Omega}_{\mathsf{K}}$ is composed of the supports of all possible K-sparse strict (local) minimizers of $\mathcal{F}_d$.*

The next propositions checks the existence of $\beta > 0$ ensuring that all the global minimizers of $\mathcal{F}_d$ are K-sparse, for some $\mathsf{K} \in \mathbb{I}_{\mathsf{M}-1}$.

Proposition 4.5. *Let $d \in \mathbb{R}^{\mathsf{M}}$. For any $\mathsf{K} \in \mathbb{I}_{\mathsf{M}-1}$, there exists $\beta_{\mathsf{K}} \geqslant 0$ such that if $\beta > \beta_{\mathsf{K}}$, then each global minimizer $\hat{u}$ of $\mathcal{F}_d$ satisfies*

$$\|\hat{u}\|_0 \leqslant \mathsf{K} \quad \text{and} \quad \sigma(\hat{u}) \in \overline{\Omega}_{\mathsf{K}} \ . \tag{44}$$

One can choose $\beta_{\mathsf{K}} = \|A\widetilde{u} - d\|^2$ where $\widetilde{u}$ solves $(\,\mathcal{P}_\omega\,)$ for some $\omega \in \Omega_{\mathsf{K}}$.

The proof is given in Appendix 8.4. The value of β_{K} in the statement is easy to compute, but in general it is not sharp[17].

5. Uniqueness of the global minimizer of $\mathcal{F}_d$. The presentation is simplified using the notation introduced next. Given a matrix $A \in \mathbb{R}^{\mathsf{M}\times\mathsf{N}}$, with any $\omega \in \Omega$ (see Definition 3.1), we associate the $\mathsf{M}\times\mathsf{M}$ matrix Π_ω that yields the orthogonal projection[18] onto the subspace spanned by the columns of A_ω:

$$\boxed{\Pi_\omega \stackrel{\text{def}}{=} A_\omega \left(A_\omega^T A_\omega\right)^{-1} A_\omega^T \ .} \tag{45}$$

For $\omega \in \Omega$, the projector in Remark 1 reads $\Pi_{\mathrm{range}(A_\omega)} = \Pi_\omega$.

Checking whether a global minimizer $\hat{u}$ of $\mathcal{F}_d$ is unique requires us to compare its value $\mathcal{F}_d(\hat{u})$ with the values $\mathcal{F}_d(\bar{u})$ of the concurrent strict minimizers $\bar{u}$. Let $\hat{u}$ be an $(\mathsf{M}-1)$-sparse strict (local) minimizer of $\mathcal{F}_d$. Then $\hat{\sigma} \stackrel{\text{def}}{=} \sigma(\hat{u}) \in \Omega$. Using Remark 1 shows that

$$\begin{aligned}\mathcal{F}_d(\hat{u}) &= \|A_{\hat{\sigma}}\hat{u}_{\hat{\sigma}} - d\|^2 + \beta\,\sharp\,\hat{\sigma} = \|\Pi_{\hat{\sigma}} d - d\|^2 + \beta\,\sharp\,\hat{\sigma} \\ &= d^T\left(I_{\mathsf{M}} - \Pi_{\hat{\sigma}}\right) d + \beta\,\sharp\,\hat{\sigma} \ .\end{aligned} \tag{46}$$

Let $\bar{u}$ be another $(\mathsf{M}-1)$-sparse strict minimizer of $\mathcal{F}_d$; set $\bar{\sigma} \stackrel{\text{def}}{=} \sigma(\bar{u})$. Then

$$\mathcal{F}_d(\hat{u}) - \mathcal{F}_d(\bar{u}) = d^T\left(\Pi_{\bar{\sigma}} - \Pi_{\hat{\sigma}}\right) d + \beta(\,\sharp\,\hat{\sigma} - \,\sharp\,\bar{\sigma}) \ .$$

If both $\hat{u}$ and $\bar{u} \neq \hat{u}$ are global minimizers of $\mathcal{F}_d$, the previous equality yields

$$\mathcal{F}_d(\hat{u}) = \mathcal{F}_d(\bar{u}) \quad \Leftrightarrow \quad d^T\left(\Pi_{\bar{\sigma}} - \Pi_{\hat{\sigma}}\right) d = -\beta(\,\sharp\,\hat{\sigma} - \,\sharp\,\bar{\sigma}) \ . \tag{47}$$

Equation (47) reveals that the uniqueness of the global minimizer of $\mathcal{F}_d$ cannot be guaranteed without suitable assumptions on A and on d.

[16] Clearly, $\overline{\Omega}_{\mathsf{M}-1} = \Omega$.

[17] For $\beta \gtrapprox \beta_{\mathsf{K}}$ the global minimizers of $\mathcal{F}_d$ might be k-sparse for $k \ll \mathsf{K}$. A sharper β_{K} can be obtained as $\beta_{\mathsf{K}} = \min_{\omega\in\Omega_{\mathsf{K}}}\left\{\|A\widetilde{u} - d\|^2 \ : \ \widetilde{u} \ \text{solve} \ (\,\mathcal{P}_\omega\,) \ \text{for} \ \ \omega \in \Omega_{\mathsf{K}}\right\}$.

[18] If $\omega = \varnothing$, we have $A_\omega \in \mathbb{R}^{\mathsf{M}\times 0}$ and so Π_ω is an $\mathsf{M}\times\mathsf{M}$ null matrix.

5.1. A generic assumption on A. We adopt an assumption on the matrix A in $\mathcal{F}_d$ in order to restrict the cases when (47) takes place for some supports $\hat\sigma \neq \bar\sigma$ obeying $\sharp\,\hat\sigma = \sharp\,\bar\sigma$.

H 1. *The matrix $A \in \mathbb{R}^{\mathsf{M}\times\mathsf{N}}$, where $\mathsf{M} < \mathsf{N}$, is such that for some given $\mathsf{K} \in \mathbb{I}_{\mathsf{M}-1}$,*

$$\boxed{r \in \mathbb{I}_{\mathsf{K}} \ \text{ and } \ (\omega, \varpi) \in (\Omega_r \times \Omega_r), \ \ \omega \neq \varpi \quad \Rightarrow \quad \Pi_\omega \neq \Pi_\varpi \ .} \tag{48}$$

Assumption H1 means that the angle (or the gap) between the equidimensional subspaces range (A_ω) and range (A_ϖ) must be nonzero [29]. For instance, if $(i, j) \in \mathbb{I}_{\mathsf{N}} \times \mathbb{I}_{\mathsf{N}}$ satisfy $i \neq j$, H1 implies that $a_i \neq \kappa\, a_j$ for any $\kappa \in \mathbb{R}\setminus\{0\}$ since $\Pi_{\{i\}} = a_i a_i^T / \|a_i\|^2$.

Checking whether H1 holds for a given matrix A requires a combinatorial search over all possible couples $(\varpi, \omega) \in (\Omega_r \times \Omega_r)$ satisfying $\varpi \neq \omega$, $\forall\, r \in \mathbb{I}_{\mathsf{K}}$. This is hard to do. Instead, we wish to know whether or not H1 is a *practical* limitation. Using some auxiliary claims, we shall show that H1 *fails only for a closed negligible subset of matrices* in the space of all $\mathsf{M}\times\mathsf{N}$ real matrices.

Lemma 5.1. *Given $r \in \mathbb{I}_{\mathsf{M}-1}$ and $\varpi \in \Omega_r$, define the following set of submatrices of A:*

$$\mathcal{H}_\varpi = \Big\{ A_\omega \ : \ \omega \in \Omega_r \ \text{ and } \ \Pi_\omega = \Pi_\varpi \Big\} \ .$$

Then $\mathcal{H}_\varpi$ belongs to an $(r \times r)$-dimensional subspace of the space of all $\mathsf{M} \times r$ matrices.

Proof. Using the fact that $\varpi \in \Omega_r$ and $\omega \in \Omega_r$, we have[19]

$$\Pi_\omega = \Pi_\varpi \quad \Leftrightarrow \quad A_\omega = \Pi_\varpi A_\omega \ . \tag{49}$$

Therefore $\mathcal{H}_\varpi$ equivalently reads

$$\mathcal{H}_\varpi = \Big\{ A_\omega \ : \ \omega \in \Omega_r \ \text{ and } \ A_\omega = \Pi_\varpi A_\omega \Big\} \ . \tag{50}$$

Let $A_\omega \in \mathcal{H}_\varpi$. Denote the columns of A_ω by $\tilde a_i$ for $i \in \mathbb{I}_r$. Then (50) yields

$$\Pi_\varpi \tilde a_i = \tilde a_i, \ \ \forall\, i \in \mathbb{I}_r \quad \Rightarrow \quad \tilde a_i \in \text{range}(A_\varpi), \ \ \forall\, i \in \mathbb{I}_r \ .$$

Hence all $\tilde a_i$, $i \in \mathbb{I}_r$, live in the r-dimensional vector subspace range(A_ϖ). All the columns of each matrix $A_\omega \in \mathcal{H}_\varpi$ belong to range(A_ϖ) as well. It follows that $\mathcal{H}_\varpi$ belongs to a (closed) subspace of dimension $r \times r$ in the space of all $\mathsf{M} \times r$ matrices, where $r \leqslant \mathsf{M} - 1$. ☐

More details on the submatrices of A living in $\mathcal{H}_\varpi$ are given next.

Remark 8. The *closed negligible* subset $\mathcal{H}_\varpi$ in Lemma 5.1 is formed from all the submatrices of A that are column equivalent to A_ϖ (see [29, p. 171]), that is,

$$A_\omega \in \mathcal{H}_\varpi \quad \Leftrightarrow \quad \exists\, P \in \mathbb{R}^{r\times r} \ \text{ such that } \ \text{rank}(P) = r \ \text{ and } \ A_\omega = A_\varpi P \ . \tag{51}$$

---

[19] Using (45), as well as the fact that $A_\omega = \Pi_\omega A_\omega$, $\forall\, \omega \in \Omega_r$, one easily derives (49) since

$$\Pi_\omega = \Pi_\varpi \ \Leftrightarrow \left\{ \begin{array}{l} A_\omega \left(A_\omega^T A_\omega\right)^{-1} A_\omega^T = \Pi_\varpi \\ A_\varpi \left(A_\varpi^T A_\varpi\right)^{-1} A_\varpi^T = \Pi_\omega \end{array} \right. \Rightarrow \left\{ \begin{array}{l} A_\omega = \Pi_\varpi A_\omega \\ A_\varpi = \Pi_\omega A_\varpi \end{array} \right. \Rightarrow \left\{ \begin{array}{l} \Pi_\omega = \Pi_\varpi \Pi_\omega \\ \Pi_\varpi = \Pi_\omega \Pi_\varpi \end{array} \right. \Rightarrow \Pi_\omega = \Pi_\varpi.$$

Observe that P has r^2 unknowns that must satisfy $\mathsf{M}r$ equations and that P must be invertible. It should be quite unlikely that such a matrix P does exist.

This remark can help to discern whether or not structured dictionaries satisfy H1.

Next we inspect the set of all matrices A failing assumption H1.

Lemma 5.2. *Consider the set $\mathcal{H}$ formed from $\mathsf{M}\times\mathsf{N}$ real matrices described next:*

$$\mathcal{H} \stackrel{\text{def}}{=} \left\{A \in \mathbb{R}^{\mathsf{M}\times\mathsf{N}} \ : \exists\, r \in \mathbb{I}_{\mathsf{M}-1},\ \exists\,(\varpi,\omega) \in \Omega_r \times \Omega_r,\ \varpi \neq \omega \ \text{ and } \ \Pi_\varpi = \Pi_\omega\right\} .$$

Then $\mathcal{H}$ belongs to a finite union of vector subspaces in $\mathbb{R}^{\mathsf{M}\times\mathsf{N}}$ whose Lebesgue measure in $\mathbb{R}^{\mathsf{M}\times\mathsf{N}}$ is null.

Proof. Let $A \in \mathcal{H}$. Then there exist at least one integer $r \in \mathbb{I}_{\mathsf{M}-1}$ and at least one pair $(\varpi,\omega) \in \Omega_r \times \Omega_r$ such that $\varpi \neq \omega$ and $\Pi_\varpi = \Pi_\omega$. Using Lemma 5.1, A contains (at least) one $\mathsf{M}\times r$ submatrix A_ϖ belonging to an $r\times r$ vector subspace in the space of all $\mathsf{M}\times r$ real matrices. Identifying A with an MN-length vector, its entries are included in a vector subspace of $\mathbb{R}^{\mathsf{MN}}$ of dimension no larger than $\mathsf{MN}-1$. The claim of the lemma is straightforward. $\square$

We can now clarify assumption H1 and show that it is really good.

Theorem 5.3. *Given an arbitrary $\mathsf{K} \in \mathbb{I}_{\mathsf{M}-1}$, consider the set of $\mathsf{M}\times\mathsf{N}$ real matrices below*

$$\mathcal{A}_{\mathsf{K}} \stackrel{\text{def}}{=} \left\{A \in \mathbb{R}^{\mathsf{M}\times\mathsf{N}} \ : \ A \ \text{ satisfies } \ \text{H1} \ \text{ for } \ \mathsf{K}\right\} .$$

Then $\mathcal{A}_{\mathsf{K}}$ contains an open and dense subset in the space of all $\mathsf{M}\times\mathsf{N}$ real-valued matrices.

Proof. The complement of $\mathcal{A}_{\mathsf{K}}$ in the space of all $\mathsf{M}\times\mathsf{N}$ real matrices reads as

$$\mathcal{A}^c_{\mathsf{K}} = \left\{A \in \mathbb{R}^{\mathsf{M}\times\mathsf{N}} \ : \ \text{H1 fails for } \ A \ \text{ and } \ \mathsf{K}\right\} .$$

It is clear that $\mathcal{A}^c_{\mathsf{K}} \subset \mathcal{H}$, where $\mathcal{H}$ is described in Lemma 5.2. By the same lemma, $\mathcal{A}^c_{\mathsf{K}}$ is included in a closed subset of vector subspaces in $\mathbb{R}^{\mathsf{M}\times\mathsf{N}}$ whose Lebesgue measure in $\mathbb{R}^{\mathsf{M}\times\mathsf{N}}$ is null. Consequently, $\mathcal{A}_{\mathsf{K}}$ satisfies the statement of the theorem. $\square$

For any $\mathsf{K} \in \mathbb{I}_{\mathsf{M}-1}$, H1 is a generic *property of all $\mathsf{M}\times\mathsf{N}$ real matrices meeting $\mathsf{M} < \mathsf{N}$.* This is the meaning of Theorem 5.3 in terms of Definition 1.2.

We can note that

$$\mathcal{A}_{\mathsf{K}+1} \subseteq \mathcal{A}_{\mathsf{K}}, \quad \forall\, \mathsf{K} \in \mathbb{I}_{\mathsf{M}-2} .$$

One can hence presume that H1 is weakened as K decreases. This issue is illustrated in section 6.

5.2. A generic assumption on d. A preliminary result is stated next.

Lemma 5.4. *Let $(\omega,\varpi) \in \overline{\Omega}_{\mathsf{K}} \times \overline{\Omega}_{\mathsf{K}}$ for $\omega \neq \varpi$ and let H1 hold for $\mathsf{K} \in \mathbb{I}_{\mathsf{M}-1}$. Given $\kappa \in \mathbb{R}$, define*

$$\mathrm{T}_\kappa \stackrel{\text{def}}{=} \{g \in \mathbb{R}^{\mathsf{M}} \ : \ g^T\left(\Pi_\omega - \Pi_\varpi\right) g = \kappa\} .$$

Then T_κ is a closed subset of $\mathbb{R}^{\mathsf{M}}$ and $\mathbb{L}^{\mathsf{M}}\left(\mathrm{T}_\kappa\right) = 0$.

Proof. Define $f : \mathbb{R}^{\mathsf{M}} \to \mathbb{R}$ by $f(g) = g^T \left(\Pi_\omega - \Pi_\varpi\right) g$. Then

$$\mathrm{T}_\kappa = \{g \in \mathbb{R}^{\mathsf{M}} \ : \ f(g) = \kappa\} \ . \tag{52}$$

Using H1, T_κ is closed in $\mathbb{R}^{\mathsf{M}}$. Set

$$\mathrm{Q} = \{g \in \mathbb{R}^{\mathsf{M}} \ : \ \nabla f(g) \neq 0\} \quad \text{and} \quad \mathrm{Q}^c = \mathbb{R}^{\mathsf{M}} \setminus \mathrm{Q}.$$

Consider an arbitrary $g \in \mathrm{T}_\kappa \cap \mathrm{Q}$. From H1, $\operatorname{rank}(\nabla f(g)) = 1$. For simplicity, assume that

$$\nabla f(g)[\mathsf{M}] = \frac{df(g)}{dg[\mathsf{M}]} \neq 0 \ .$$

By the implicit functions theorem, there are open neighborhoods $\mathcal{O}_g \subset \mathrm{Q} \subset \mathbb{R}^{\mathsf{M}}$ and $\mathcal{V} \subset \mathbb{R}^{\mathsf{M}-1}$ of g and $g_{\mathbb{I}_{\mathsf{M}-1}}$, respectively, and a unique $\mathcal{C}^1$-function $h_g : \mathcal{V} \to \mathbb{R}$ with ∇h_g bounded, such that

$$\gamma = (\gamma_{\mathbb{I}_{\mathsf{M}-1}}, \gamma[\mathsf{M}]) \in \mathcal{O}_g \ \text{ and } \ f(\gamma) = \kappa \quad \Leftrightarrow \quad \gamma_{\mathbb{I}_{\mathsf{M}-1}} \in \mathcal{V} \ \text{ and } \ \gamma[\mathsf{M}] = h_g(\gamma_{\mathbb{I}_{\mathsf{M}-1}}) \ . \tag{53}$$

From (52) and (53) it follows that[20]

$$\mathcal{O}_g \cap \mathrm{T}_k = \psi^g \left(\mathcal{O}_g \cap (\mathbb{R}^{\mathsf{M}-1} \times \{0\})\right) ,$$

where ψ^g is a diffeomorphism on $\mathcal{O}_g$ given by

$$\psi^g_i(\gamma) = \gamma[i], \quad 1 \leqslant i \leqslant \mathsf{M}-1 \quad \text{and} \quad \psi^g_{\mathsf{M}}(\gamma) = h_g(\gamma_{\mathbb{I}_{\mathsf{M}-1}}) + \gamma[\mathsf{M}] \ .$$

Since $\mathbb{L}^{\mathsf{M}}\left(\mathcal{O}_g \cap (\mathbb{R}^{\mathsf{M}-1} \times \{0\})\right) = 0$ and $\nabla\psi^g$ is bounded on $\mathcal{O}_g$, it follows from [37, Lemma 7.25] that[21] $\mathbb{L}^{\mathsf{M}}\left(\mathcal{V}_g \cap \mathrm{T}_k\right) = 0$. We have thus obtained that

$$S \subset Q \ \text{ and } \ S \text{ bounded} \quad \Rightarrow \quad \mathbb{L}^{\mathsf{M}}(S \cap \mathrm{T}_k) = 0 \ . \tag{54}$$

Using that every open subset of $\mathbb{R}^{\mathsf{M}}$ can be written as a countable union[22] of cubes in $\mathbb{R}^{\mathsf{M}}$ [36, 16, 38], the result in (54) entails that $\mathbb{L}^{\mathsf{M}}(\mathrm{T}_\kappa \cap Q) = 0$.

Next, $Q^c = \ker\left(\Pi_\omega - \Pi_\varpi\right)$. By H1, $\dim\ker\left(\Pi_\omega - \Pi_\varpi\right) \leqslant \mathsf{M}-1$. Hence $\mathbb{L}^{\mathsf{M}}(\mathrm{T}_\kappa \cap Q^c) = 0$.

The proof follows from the equality $\mathbb{L}^{\mathsf{M}}(\mathrm{T}_\kappa) = \mathbb{L}^{\mathsf{M}}(\mathrm{T}_\kappa \cap Q) + \mathbb{L}^{\mathsf{M}}(\mathrm{T}_\kappa \cap Q^c)$. ☐

We exhibit a *closed negligible* subset of data in $\mathbb{R}^{\mathsf{M}}$ that can still meet the equality in (47).

Proposition 5.5. *For $\beta > 0$ and $\mathsf{K} \in \mathbb{I}_{\mathsf{M}-1}$, put*

$$\Sigma_{\mathsf{K}} \stackrel{\text{def}}{=} \bigcup_{n=-\mathsf{K}}^{\mathsf{K}} \ \bigcup_{\omega \in \overline{\Omega}_{\mathsf{K}}} \ \bigcup_{\varpi \in \overline{\Omega}_{\mathsf{K}}} \left\{ g \in \mathbb{R}^{\mathsf{M}} \ : \ \omega \neq \varpi \ \text{ and } \ g^T \Big(\Pi_\omega - \Pi_\varpi\Big) g = n\beta \right\} \ , \tag{55}$$

where $\overline{\Omega}_{\mathsf{K}}$ is given in (43). Let H1 hold for K. Then Σ_{K} is closed in $\mathbb{R}^{\mathsf{M}}$ and $\mathbb{L}^{\mathsf{M}}\left(\Sigma_{\mathsf{K}}\right) = 0$.

[20] From (53), $\mathcal{V}$ is the restriction of $\mathcal{O}_g$ to $\mathbb{R}^{\mathsf{M}-1}$.

[21] The same result follows from the change-of-variables theorem for the Lebesgue integral, see e.g. [37].

[22] From (54), adjacent cubes can also intersect in our case.

Proof. For some $n \in \{-\mathsf{K}, \cdots, \mathsf{K}\}$ and $(\omega, \varpi) \in (\overline{\Omega}_\mathsf{K} \times \overline{\Omega}_\mathsf{K})$ such that $\omega \neq \varpi$, put

$$\Sigma \stackrel{\text{def}}{=} \left\{ g \in \mathbb{R}^\mathsf{M} : g^T \Big(\Pi_\omega - \Pi_\varpi \Big) g = n\beta \right\} .$$

If $\sharp\,\omega \neq \sharp\,\varpi$, then $\operatorname{rank}\big(\Pi_\omega - \Pi_\varpi\big) \geqslant 1$. If $\sharp\,\omega = \sharp\,\varpi$, H1 guarantees that $\operatorname{rank}\big(\Pi_\omega - \Pi_\varpi\big) \geqslant 1$, yet again. The number $n\beta \in \mathbb{R}$ is given. According to Lemma 5.4, Σ is a closed subset of $\mathbb{R}^\mathsf{M}$ and $\mathbb{L}^\mathsf{M}(\Sigma) = 0$. The conclusion follows from the fact that Σ_K is a finite union of subsets like Σ. $\square$

We assume hereafter that if H1 holds for some $\mathsf{K} \in \mathbb{I}_{\mathsf{M}-1}$, data d satisfy

$$d \in \{ g \in \mathbb{R}^\mathsf{M} \ : \ g \notin \Sigma_\mathsf{K} \} = \mathbb{R}^\mathsf{M} \setminus \Sigma_\mathsf{K} .$$

5.3. The unique global minimizer of $\mathcal{F}_d$ is K-sparse for $\mathsf{K} \leqslant (\mathsf{M}-1)$.

We are looking for guarantees that $\mathcal{F}_d$ has a *unique* global minimizer $\hat{u}$ obeying

$$\|\hat{u}\|_0 \leqslant \mathsf{K} \quad \text{for some fixed} \quad \mathsf{K} \in \mathbb{I}_{\mathsf{M}-1} .$$

This is the aim of the next theorem.

Theorem 5.6. *Given $\mathsf{K} \in \mathbb{I}_{\mathsf{M}-1}$, let H1 hold for K, $\beta > \beta_\mathsf{K}$ where β_K meets Proposition 4.5 and $\Sigma_\mathsf{K} \subset \mathbb{R}^\mathsf{M}$ reads as in (55). Consider that*

$$d \in \mathbb{R}^\mathsf{M} \setminus \Sigma_\mathsf{K} .$$

Then

(i) *the set $\mathbb{R}^\mathsf{M} \setminus \Sigma_\mathsf{K}$ is open and dense in $\mathbb{R}^\mathsf{M}$;*

(ii) *$\mathcal{F}_d$ has a* unique global minimizer *$\hat{u}$, and $\|\hat{u}\|_0 \leqslant \mathsf{K}$.*

Proof. Statement (i) follows from Proposition 5.5.

Since $\beta > \beta_\mathsf{K}$, all global minimizers of $\mathcal{F}_d$ have their support in $\overline{\Omega}_\mathsf{K}$ (Proposition 4.5). Using the fact that $d \in \mathbb{R}^\mathsf{M} \setminus \Sigma_\mathsf{K}$, the definition of Σ_K in (55) shows that

$$-\mathsf{K} \leqslant n \leqslant \mathsf{K} \ \text{ and } \ (\omega, \varpi) \in (\overline{\Omega}_\mathsf{K} \times \overline{\Omega}_\mathsf{K}),\ \omega \neq \varpi \quad \Rightarrow \quad d^T \Big(\Pi_\omega - \Pi_\varpi \Big) d \neq n\beta . \tag{56}$$

The proof is conducted by contradiction. Let $\hat{u}$ and $\bar{u} \neq \hat{u}$ be two global minimizers of $\mathcal{F}_d$. Then

$$\hat{\sigma} \stackrel{\text{def}}{=} \sigma(\hat{u}) \in \overline{\Omega}_\mathsf{K} \quad \text{and} \quad \bar{\sigma} \stackrel{\text{def}}{=} \sigma(\bar{u}) \in \overline{\Omega}_\mathsf{K} ,$$

and $\hat{\sigma} \neq \bar{\sigma}$. By $\mathcal{F}_d(\hat{u}) = \mathcal{F}_d(\bar{u})$, (47) yields

$$d^T \left(\Pi_{\hat{\sigma}} - \Pi_{\bar{\sigma}} \right) d = \beta(\sharp\,\hat{\sigma} - \sharp\,\bar{\sigma}) . \tag{57}$$

An enumeration of all possible values of $\sharp\,\hat{\sigma} - \sharp\,\bar{\sigma}$ shows that

$$\beta\,(\sharp\,\hat{\sigma} - \sharp\,\bar{\sigma}) = n\beta \quad \text{for some} \quad n \in \{-\mathsf{K}, \cdots, \mathsf{K}\} .$$

Inserting this equation into (57) leads to

$$d^T\left(\Pi_{\hat\sigma}-\Pi_{\bar\sigma}\right)d = n\beta \quad \text{for some} \quad n\in\{-\mathsf{K},\cdots,\mathsf{K}\}\ .$$

The last result contradicts (56); hence it violates the assumptions H1 and $d\in\mathbb{R}^{\mathsf{M}}\setminus\Sigma_{\mathsf{K}}$. Consequently, $\mathcal{F}_d$ cannot have two global minimizers. Since $\mathcal{F}_d$ always has global minimizers (Theorem 4.4(i)), it follows that $\mathcal{F}_d$ has a unique global minimizer, say $\hat u$. And $\|\hat u\|_0\leqslant\mathsf{K}$ because $\sigma(\hat u)\in\overline{\Omega}_{\mathsf{K}}$. □

For $\beta>\beta_{\mathsf{K}}$, the objective $\mathcal{F}_d$ in (1) *has a unique global minimizer and it is* K*-sparse for* $\mathsf{K}\leqslant\mathsf{M}-1$. *For all* $\mathsf{K}\in\mathbb{I}_{\mathsf{M}-1}$, *the claim holds true in a generic sense.* This is the message of Theorem 5.6 using Definition 1.2.

6. Numerical illustrations.

6.1. On assumption H1. Assumption H1 requires that $\Pi_\omega\neq\Pi_\varpi$ when $(\omega,\varpi)\in\Omega_r\times\Omega_r$, $\omega\neq\varpi$ for all $r\leqslant\mathsf{K}\in\mathbb{I}_{\mathsf{M}-1}$. From a practical viewpoint, the magnitude of $(\Pi_\omega-\Pi_\varpi)$ should be discernible. One way to assess the viability of H1 for a matrix A and $\mathsf{K}\in\mathbb{I}_{\mathsf{M}-1}$ is to calculate

$$\xi_{\mathsf{K}}(A)\overset{\text{def}}{=}\min_{r\in\mathbb{I}_{\mathsf{K}}}\ \mu_r(A)\ , \tag{58}$$
$$\text{where}\qquad \mu_r(A)=\min_{\substack{(\omega,\varpi)\,\in\,\Omega_r\times\Omega_r\\ \omega\neq\varpi}}\ \|\Pi_\omega-\Pi_\varpi\|_2\,,\qquad\forall\, r\in\mathbb{I}_{\mathsf{K}}\ .$$

In fact, $\|\Pi_\omega-\Pi_\varpi\|_2=\sin(\theta)$, where $\theta\in[0,\pi/2]$ is the maximum angle between range$\,(A_\omega)$ and range$\,(A_\varpi)$; see [29, p. 456]. These subspaces have the same dimension and $\Pi_\omega\neq\Pi_\varpi$ when $(\omega,\varpi)\in\Omega_r\times\Omega_r$, $\omega\neq\varpi$ and $r\in\mathbb{I}_{\mathsf{K}}$, hence $\theta\in(0,\pi/2]$. Consequently,

$$\text{H1 holds for }\ \mathsf{K}\in\mathbb{I}_{\mathsf{M}-1}\quad\Rightarrow\quad\mu_r(A)\in(0,1]\ \ \forall\, r\in\mathbb{I}_{\mathsf{K}}\quad\Rightarrow\quad\xi_{\mathsf{K}}(A)\in(0,1]\ .$$

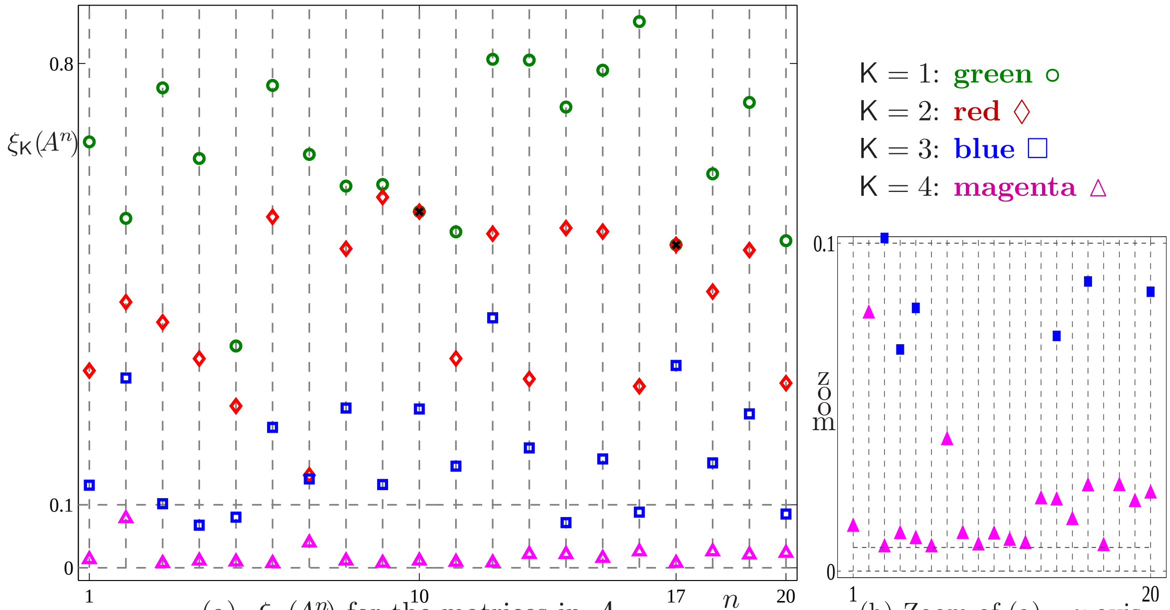
According to (58), we have $\xi_{\mathsf{K}}\geqslant\xi_{\mathsf{K}+1}$, $\forall\,\mathsf{K}\in\mathbb{I}_{\mathsf{M}-2}$. Our guess that *assumption H1 is lightened when* K *decreases* (see the comments following the proof of Theorem 5.3) means that

$$\xi_1(A)>\cdots>\xi_{\mathsf{M}-1}(A)\ . \tag{59}$$

We provide numerical tests on two subsets of real-valued random matrices for $\mathsf{M}=5$ and $\mathsf{N}=10$, denoted by $\mathcal{A}^N_{20}$ and $\mathcal{A}^U_{1000}$. The values of $\xi_{\mathsf{K}}(\cdot)$, $\mathsf{K}\in\mathbb{I}_{\mathsf{M}-1}=\mathbb{I}_4$, for every matrix in $\mathcal{A}^N_{20}$ and in $\mathcal{A}^U_{1000}$, were calculated using an *exhaustive combinatorial search*. *All tested matrices satisfy assumption H1*, which confirms Theorem 5.3 and its consequences. In order to evaluate the extent of H1, we computed the *worst* and the *best* values of $\xi_{\mathsf{K}}(\cdot)$ over these sets:

$$\begin{cases}\xi_{\mathsf{K}}^{worst} &= \displaystyle\min_{A\in\mathcal{A}}\xi_{\mathsf{K}}(A)\\[2ex] \xi_{\mathsf{K}}^{best} &= \displaystyle\max_{A\in\mathcal{A}}\xi_{\mathsf{K}}(A)\end{cases}\qquad\forall\,\mathsf{K}\in\mathbb{I}_{\mathsf{M}-1}\ ,\quad\mathcal{A}\in\left\{\mathcal{A}^N_{20},\ \mathcal{A}^U_{1000}\right\}\ . \tag{60}$$

Set $\mathcal{A}^N_{20}$. This set was formed from 20 matrices A^n, $n \in \mathbb{I}_{20}$ of size 5×10. The components of each matrix A^n were independent and uniformly drawn from the standard normal distribution with mean zero and variance one. The values of $\xi_{\mathsf{K}}(\cdot)$ are depicted in Fig. 6.1. We have[23] $\xi_1(A^{10}) = \xi_2(A^{10})$ and $\xi_1(A^{17}) = \xi_2(A^{17})$. In all other cases (59) is satisfied. Fig. 6.1 clearly shows that $\xi_{\mathsf{K}}(\cdot)$ increases as K decreases (from $\mathsf{M}-1$ to 1).



(a) $\xi_{\mathsf{K}}(A^n)$ for the matrices in $\mathcal{A}_{20}$ (b) Zoom of (a) – y-axis

Figure 6.1. *x-axis: the list of the 20 random matrices in* $\mathcal{A}^N_{20}$. *(a) y-axis: the value* $\xi_{\mathsf{K}}(A^n)$ *according to* (58) *for all* $\mathsf{K} \in \mathbb{I}_{\mathsf{M}-1}$ *and for all* $n \in \mathbb{I}_{20}$. *The plot in (b) is a zoom of (a) along the y-axis.*

The worst and the best values of $\xi_{\mathsf{K}}(\cdot)$ over the whole set $\mathcal{A}^N_{20}$ are displayed in Table 6.1.

Table 6.1

The worst and the best values of $\xi_{\mathsf{K}}(A)$, *for* $\mathsf{K} \in \mathbb{I}_{\mathsf{M}-1}$, *over the set* $\mathcal{A}^N_{20}$, *see* (60).

	$\mathsf{K}=1$	$\mathsf{K}=2$	$\mathsf{K}=3$	$\mathsf{K}=4$
ξ_{K}^{worst}	0.3519	0.1467	0.0676	0.0072
ξ_{K}^{best}	0.8666	0.5881	0.3966	0.0785

Set $\mathcal{A}^U_{1000}$. The set $\mathcal{A}^U_{1000}$ was composed of one thousand 5×10 matrices A^n, $n \in \mathbb{I}_{1000}$. The entries of each matrix A^n were independent and uniformly sampled on $[-1, 1]$. The obtained values for ξ_{K}^{worst} and ξ_{K}^{best}, calculated according to (60), are shown in Table 6.2.

For $\mathsf{K} \in \mathbb{I}_3$, the *best* values of $\xi_{\mathsf{K}}(\cdot)$ were obtained for the same matrix, A^{964}. Note that $\xi_4(A^{964}) = 0.0425 \gg \xi_4^{worst}$. The *worst* values in Table 6.2 are smaller than those in Table 6.1, while the *best* values in Table 6.2 are larger than those in Table 6.1; one credible reason is that $\mathcal{A}^U_{1000}$ is much larger than $\mathcal{A}^N_{20}$.

[23] This is why on the figure, in columns 10 and 17, the green "○" and the red "◊" overlap.

Table 6.2
The worst and the best values of $\xi_{\mathsf{K}}(A)$, for $\mathsf{K} \in \mathbb{I}_{\mathsf{M}-1}$, over the set $\mathcal{A}^U_{1000}$.

	$\mathsf{K}=1$	$\mathsf{K}=2$	$\mathsf{K}=3$	$\mathsf{K}=4$
ξ_{K}^{worst}	0.1085	0.0235	0.0045	0.0001
ξ_{K}^{best}	0.9526	0.8625	0.5379	0.1152

Table 6.3
Percentage of the cases in $\mathcal{A}^U_{1000}$ when (59) fails to hold.

	$\xi_1(A^n)=\xi_2(A^n)$	$\xi_2(A^n)=\xi_3(A^n)$	$\xi_3(A^n)=\xi_4(A^n)$
occurrences $\{n\}$	5 %	1.6 %	0.1 %

Overall, (59) is satisfied on $\mathcal{A}^U_{1000}$—the percentages in Table 6.3 are pretty small. All three tables and Figure 6.1 agree with our guess that H1 is more viable for smaller values of K.

Based on the magnitudes for ξ_{K}^{best} in Tables 6.1 and 6.2, one can expect that there are some classes of matrices (random or not) that fit H1 for larger values of $\xi_{\mathsf{K}}(\cdot)$.

6.2. On the global minimizers of $\mathcal{F}_d$. Here we summarize the outcome of a series of experiments corresponding to several matrices $A \in \mathbb{R}^{\mathsf{M}\times\mathsf{N}}$ where $\mathsf{M}=5$ and $\mathsf{N}=10$, satisfying H1 for $\mathsf{K}=\mathsf{M}-1$, different original vectors $\ddot{u} \in \mathbb{R}^{\mathsf{N}}$ and data samples $d = A\ddot{u} + \text{noise}$, for various values of $\beta > 0$. In each experiment, we computed the complete list of all different strict (local) minimizers of $\mathcal{F}_d$, say $\left(\hat{u}^i\right)_{i=1}^n$. Then the sequence of values $\left(\mathcal{F}_d(\hat{u}^i)\right)_{i=1}^n$ was sorted in increasing order, $\mathcal{F}_d\left(\hat{u}^{i_1}\right) \leqslant \mathcal{F}_d\left(\hat{u}^{i_2}\right) \leqslant \cdots \leqslant \mathcal{F}_d\left(\hat{u}^{i_n}\right)$. A global minimizer $\hat{u}^{i_1}$ is unique provided that $\mathcal{F}_d\left(\hat{u}^{i_1}\right) < \mathcal{F}_d\left(\hat{u}^{i_2}\right)$. In order to discard numerical errors, we also checked whether $\left|\mathcal{F}_d\left(\hat{u}^{i_1}\right) - \mathcal{F}_d\left(\hat{u}^{i_2}\right)\right|$ is easy to detect.

In all experiments we carried out, the following facts were observed:

- *The global minimizer of $\mathcal{F}_d$ was unique—manifestly data d never did belong to the* closed negligible *subset Σ_{K} in Proposition 5.5. This confirms Theorem 5.6.*
- *The global minimizers of $\mathcal{F}_d$ remained unchanged under large variations of β.*
- *The necessary condition for a global minimizer in Proposition 4.1 was met.*

Next we present in detail two of these experiments where $\mathcal{F}_d$ is defined using

$$(61)\quad A = \begin{bmatrix} 7 & 2 & 4 & 9 & 0 & 3 & 3 & 6 & 6 & 7 \\ 3 & 4 & 9 & 3 & 3 & 9 & 1 & 3 & 1 & 5 \\ 5 & 4 & 2 & 4 & 0 & 7 & 1 & 9 & 2 & 9 \\ 8 & 4 & 0 & 9 & 6 & 0 & 4 & 2 & 3 & 7 \\ 6 & 3 & 6 & 5 & 0 & 9 & 0 & 0 & 3 & 8 \end{bmatrix} \qquad \begin{array}{l} d = A\ddot{u} + n\ , \\ \\ \quad \text{where } n \text{ is noise and} \\ \\ \ddot{u} = \big(0,\ 1,\ 8,\ 0,\ 3,\ 0,\ 0,\ 0,\ 0,\ 9\big)^T. \end{array}$$

Only integers appear in (61) for better readability. We have $\operatorname{rank}(A) = \mathsf{M} = 5$. An exhaustive combinatorial test shows that the arbitrary matrix A in (61) satisfies H1 for $\mathsf{K}=\mathsf{M}-1$. The values of $\xi_{\mathsf{K}}(A)$ are seen in Table 6.4. One notes that $\mu_2(A) > \mu_1(A)$; hence $\xi_1(A) = \xi_2(A)$.

One *expects* (at least when data are noise-free) that the global minimizer $\hat{u}$ of $\mathcal{F}_d$ obeys $\hat{\sigma} \subseteq \sigma(\ddot{u})$, where $\ddot{u}$ is the original in (61), and that the vanished entries of $\hat{u}$ correspond to the least entries of $\ddot{u}$. This inclusion provides a partial way to rate the quality of the solution provided by a global minimizer $\hat{u}$ of $\mathcal{F}_d$.

Table 6.4

The values of $\xi_{\mathsf{K}}(A)$ and $\mu_{\mathsf{K}}(A)$, $\forall\, \mathsf{K} \in \mathbb{I}_{\mathsf{M}-1}$, for the matrix A in (61).

	$\mathsf{K}=1$	$\mathsf{K}=2$	$\mathsf{K}=3$	$\mathsf{K}=4$
$\xi_{\mathsf{K}}(A)$	0.2737	0.2737	0.2008	0.0564
$\mu_{\mathsf{K}}(A)$	0.2737	0.2799	0.2008	0.0564

The experiments described hereafter correspond to two data samples relevant to (61)—without and with noise—and to several values of $\beta > 0$.

Noise-free data. The noise-free data in (61) read as:

$$d = A\ddot{u} = \left(\ 97\,,\quad 130\,,\quad 101\,,\quad 85\,,\quad 123\ \right)^T. \tag{62}$$

For different values of β, the global minimizer $\hat{u}$ is given in Table 6.5. Since $\sigma(\ddot{u}) \in \Omega$ and

Table 6.5

The global minimizer $\hat{u}$ of $\mathcal{F}_d$ and its value $\mathcal{F}_d(\hat{u})$ for the noise-free data d in (62) for different values of β. Last row: the original $\ddot{u}$ in (61).

β	The global minimizer $\hat{u}$ of $\mathcal{F}_d$ *(row vector)*										$\Vert\hat{u}\Vert_0$	$\mathcal{F}_d(\hat{u})$
1	0	**1**	**8**	0	**3**	0	0	0	0	**9**	4	4
10^2	0	0	**8.12**	0	**3.31**	0	0	0	0	**9.33**	3	301.52
10^3	0	0	0	0	0	**12.58**	**20.28**	0	0	0	2	2179.3
10^4	0	**29.95**	0	0	0	0	0	0	0	0	1	14144
$7{\cdot}10^4$	0	0	0	0	0	0	0	0	0	0	0	58864
$\ddot{u}\ \ =$	0	**1**	**8**	0	**3**	0	0	0	0	**9**		

the data are noise-free, $\mathcal{F}_d$ does not have global minimizers with $\Vert\hat{u}\Vert_0 = 5$. Actually, applying Proposition 4.5 for $\widetilde{u} = \ddot{u}$ yields $\beta_{\mathsf{M}-1} = 0$, hence for any $\beta > 0$ all global minimizers of $\mathcal{F}_d$ have a support in $\Omega = \overline{\Omega}_{\mathsf{M}-1}$ (see Definition 3.1 and (43)). The global minimizer $\hat{u}$ for $\beta = 1$ meets $\hat{u} = \ddot{u}$. For $\beta = 100$, the global minimizer $\hat{u}$ obeys $\hat{\sigma} \overset{\text{def}}{=} \sigma(\hat{u}) = \{3, 5, 10\} \subsetneq \sigma(\ddot{u})$ and $\Vert\hat{u}\Vert_0 = 3$—the least nonzero entry of the original $\ddot{u}$ is canceled, which is reasonable. The global minimizers corresponding to $\beta \geqslant 300$ are meaningless. We could not find any positive value of β giving better 2-sparse global minimizers. Recalling that data are noise-free, this confirms Remark 3: the global minimizers of $\mathcal{F}_d$ realize a only *pseudo*-hard thresholding. For $\beta \geqslant 7 \cdot 10^4 > \Vert d\Vert^2$, the global minimizer of $\mathcal{F}_d$ is $\hat{u} = 0$ which confirms Remark 7.

Noisy data. Now we consider *noisy data* in (61) for

$$n = \left(\ 4\,,\quad -1\,,\quad 2\,,\quad -3\,,\quad 5\ \right)^T. \tag{63}$$

This arbitrary noise yields a signal-to-noise ratio[24] (SNR) equal to 14.07 dB. If $\beta \leqslant 0.04$, $\mathcal{F}_d$ has 252 different strict global minimizers $\hat{u}$ obeying $\Vert\hat{u}\Vert_0 = \mathsf{M}$ and $\mathcal{F}_d(\hat{u}) = \beta\mathsf{M}$ (recall Proposition 3.9). For $\beta \geqslant 0.05$, the global minimizer $\hat{u}$ of $\mathcal{F}_d$ is unique and satisfies $\sigma(\hat{u}) \in \Omega$. It is given in Table 6.6 for several values of $\beta \geqslant 0.05$. For $\beta = 1$, the global minimizer is

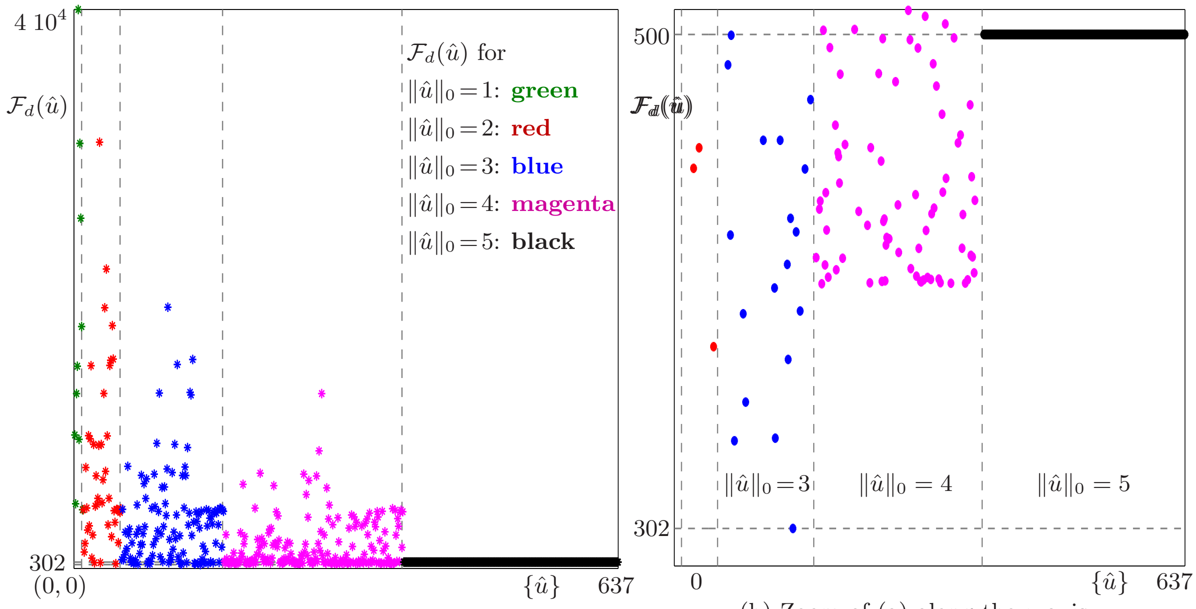
[24] Let us denote $\ddot{d} = A\ddot{u}$ and $d = \ddot{d} + n$. The SNR reads [41] $\mathrm{SNR}(\ddot{d}, d) = 10\,\log_{10} \dfrac{\sum_{i=1}^{\mathsf{M}}\left(\ddot{d}[i] - \frac{1}{\mathsf{M}}\sum_{i=1}^{\mathsf{M}} \ddot{d}[i]\right)^2}{\sum_{i=1}^{\mathsf{M}}\left(d[i] - \ddot{d}[i]\right)^2}$.

Table 6.6

The global minimizer $\hat{u}$ of $\mathcal{F}_d$ and its value $\mathcal{F}_d(\hat{u})$ for noisy data given by (61) and (63), for different values of β. Last row: the original $\ddot{u}$.

β	The global minimizer $\hat{u}$ of $\mathcal{F}_d$ *(row vector)*										$\|\hat{u}\|_0$	$\mathcal{F}_d(\hat{u})$
1	0	**6.02**	**2.66**	**6.43**	0	**6.85**	0	0	0	0	4	4.0436
10^2	0	0	**8.23**	0	**2.3**	0	0	0	0	**9.71**	3	301.94
10^3	0	0	**8.14**	0	0	0	0	0	0	**10.25**	2	2174.8
10^4	0	0	0	0	0	0	0	0	0	**14.47**	1	14473
$7{\cdot}10^4$	0	0	0	0	0	0	0	0	0	0	0	60559
$\ddot{u}$ =	0	**1**	**8**	0	**3**	0	0	0	0	**9**		

meaningless. We could not find any positive value of β yielding a better global minimizer with a 4-length support. For the other values of β, the global minimizer $\hat{u}$ meets $\hat{\sigma} \overset{\text{def}}{=} \sigma(\hat{u}) \subsetneqq \sigma(\ddot{u})$, and its vanished entries correspond to the least entries in the original $\ddot{u}$. For $\beta = 100$, the global minimizer seems to furnish a good approximation to $\ddot{u}$. Observe that the last entry of the global minimizer $\hat{u}[10]$, corresponding to the largest magnitude in $\ddot{u}$, freely increases when β increases from 10^2 to 10^4. We tested a tight sequence of intermediate values of β without finding better results. Yet again, $\beta \geqslant 7 \cdot 10^4 > \|d\|^2$ leads to a unique null global minimizer (see Remark 7).



(a) The values of all strict (local) minimizers of $\mathcal{F}_d$

(b) Zoom of (a) along the y-axis

Figure 6.2. *All* **638** *strict (local) minima of $\mathcal{F}_d$ in (61) for $\beta = 100$ and data d corrupted with the arbitrary noise in (63). The x-axis lists all strict (local) minimizers $\{\hat{u}\}$ of $\mathcal{F}_d$ sorted according to their ℓ_0-norm $\|\hat{u}\|_0$ in increasing order. (a) The y-axis shows the value $\mathcal{F}_d(\hat{u})$ of these minimizers marked with a star. The value of $\mathcal{F}_d$ for $\hat{u} = 0$ is not shown because it is too large ($\mathcal{F}_d(0) = 60559 = \|d\|^2$). (b) A zoom of (a) along the y-axis. It clearly shows that $\mathcal{F}_d$ has a very recognizable unique global minimizer*

Figure 6.2 shows the value $\mathcal{F}_d(\hat{u})$ of all the strict local minimizers of $\mathcal{F}_d$ for $\beta = 100$. In the zoom in Figure 6.2(b) it is easily seen that the global minimizer is unique (remember Theorem 5.6). It obeys $\|\hat{u}\|_0 = 3$ and $\mathcal{F}_d(\hat{u}) = 301.94$. One observes that $\mathcal{F}_d$ has $252 = \sharp\,\Omega_{\mathsf{M}}$ different strict local minimizers $\hat{u}$ with $\|\hat{u}\|_0 = 5 = \mathsf{M}$ and $\mathcal{F}_d(\hat{u}) = 500 = \beta\mathsf{M}$. This confirms Proposition 3.9—obviously d does not belong to the closed negligible subset Q_{M} described in the proposition.

7. Conclusions and perspectives.

We provided a detailed analysis of the (local and global) minimizers of a regularized objective $\mathcal{F}_d$ composed of a quadratic data fidelity term and an ℓ_0 penalty weighted by a parameter $\beta > 0$, as given in (1). We exhibited easy necessary and sufficient conditions ensuring that a (local) minimizer $\hat{u}$ of $\mathcal{F}_d$ is strict (Theorem 3.2). The global minimizers of $\mathcal{F}_d$ (whose existence was proved) were shown to be strict as well (Theorem 4.4). Under very mild conditions, $\mathcal{F}_d$ was shown to have a unique global minimizer (Theorem 5.6). Other interesting results were listed in the abstract. Below we pose some perspectives and open questions raised by this work.

- The relationship between the value of the regularization parameter β and the sparsity of the global minimizers of $\mathcal{F}_d$ (Proposition 4.5) can be improved.
- The *generic* linearity in data d of each strict (local) minimizer of $\mathcal{F}_d$ (subsection 3.2) should be exploited to better characterize the global minimizers of $\mathcal{F}_d$.
- Is there a simple way to check whether assumption H1 is satisfied by a given matrix $A \in \mathbb{R}^{\mathsf{M}\times\mathsf{N}}$ when N and $\mathsf{M} < \mathsf{N}$ are large? Remark 8 and in particular (51) could help to discard some nonrandom matrices. Conversely, one can ask whether there is a systematic way to construct matrices A that satisfy H1.
  An alternative would be to exhibit families of matrices that satisfy H1 for large values of $\xi_{\mathsf{K}}(\cdot)$, where the latter quantifiers are defined in equation (58).
- A proper adaptation of the results to matrices A and data d with complex entries should not present inherent difficulties.
- The theory developed here can be extended to MAP energies of the form evoked in (5). This is important for the imaging applications mentioned there.
- Based on Corollary 2.5, and Remarks 3 and 6, and the numerical tests in subsection 6.2, one is justified in asking for conditions ensuring that the global minimizers of $\mathcal{F}_d$ perform a valid work. Given the high quality of the numerical results provided in many papers (see e.g., [33, 34]), the question deserves attention.

There exist numerous algorithms aimed at approximating a (local) minimizer of $\mathcal{F}_d$. As a by-product of our research, we obtained simple rules to verify whether or not an algorithm could find

- a (local) minimizer $\hat{u}$ of $\mathcal{F}_d$—by checking whether $\hat{u}$ satisfies (26) in Corollary 2.5;
- and whether this local minimizer is strict by testing whether the submatrix whose columns are indexed by the support of $\hat{u}$ (i.e., $A_{\sigma(\hat{u})}$) has full column rank (Theorem 3.2).

Some properties of the minimizers of $\mathcal{F}_d$ given in this work can be inserted in numerical schemes in order to quickly escape from shallow local minimizers.

Many existing numerical methods involve a studious choice of the regularization parameter β, and some of them are proved to converge to a local minimizer of $\mathcal{F}_d$. *We have seen that*

finding a (strict or nonstrict) local minimizer of $\mathcal{F}_d$ is easy and that it is independent of the value of β (Corollaries 2.5 and 3.3). It is therefore obscure what meaning to attach to "choosing a good β and proving (local) convergence". Other successful algorithms are not guaranteed to converge to a local minimizer of $\mathcal{F}_d$. Whenever algorithms do a good job, the choice of β, the assumptions on A and on $\|\hat{u}\|_0$, and the iterative scheme and its initialization *obviously* provide a tool for selecting a meaningful solution by minimizing $\mathcal{F}_d$. *There is a theoretical gap that needs clarification.*

The connection between the existing algorithms and the description of the minimizers exposed in this paper deserves deep exploration. What conditions ensure that an algorithm minimizing $\mathcal{F}_d$ yields meaningful solutions? Clearly, showing local convergence does not answer this important question.

One can expect such research to give rise to innovative and more efficient algorithms enabling one to compute relevant solutions by minimizing the tricky objective $\mathcal{F}_d$.

8. Appendix.

8.1. Proof of Lemma 2.1. Since $\hat{u} \neq 0$, the definition of $\hat{\sigma}$ shows that $\min_{i\in\hat{\sigma}} \big|\, \hat{u}[i] \,\big| > 0$. Then ρ in (19) fulfills $\rho > 0$.

(i). Since $\sharp\, \hat{\sigma} \geqslant 1$, we have

$$
\begin{aligned}
i \in \hat{\sigma}\ ,\quad v \in \mathrm{B}_\infty(0,\rho) \quad &\Rightarrow \quad \max_{j\in\hat{\sigma}} \big|\, v[j] \,\big| < \rho \\
&\Rightarrow \quad \max_{j\in\hat{\sigma}} \big|\, v[j] \,\big| < \min_{j\in\hat{\sigma}} \big|\, \hat{u}[j] \,\big| \\
&\Rightarrow \quad |\hat{u}[i] + v[i]| \geqslant |\hat{u}[i]| - |v[i]| \\
&\qquad\ \geqslant \min_{j\in\hat{\sigma}} |\hat{u}[j]| - \max_{j\in\hat{\sigma}} |v[j]| \geqslant \rho - \max_{j\in\hat{\sigma}} |v[j]| > 0 \\
&\Rightarrow \quad \hat{u}[i] + v[i] \neq 0 \\
\Big[\text{by } (2)\,\Big] \quad &\Rightarrow \quad \phi\,(\hat{u}[i] + v[i]) = \phi\,(\hat{u}[i]) = 1\ .
\end{aligned}
\tag{64}
$$

If $\hat{\sigma}^c = \varnothing$ the result is proved. Let $\hat{\sigma}^c \neq \varnothing$. Then $\hat{u}[i] = 0 = \phi\,(\hat{u}[i])$, $\forall\, i \in \hat{\sigma}^c$. Inserting this and (64) into

$$
\sum_{i\in\mathbb{I}_\mathsf{N}} \phi\big(\, \hat{u}[i] + v[i] \,\big) = \sum_{i\in\hat{\sigma}} \phi\big(\, \hat{u}[i] + v[i] \,\big) + \sum_{i\in\hat{\sigma}^c} \phi\big(\, \hat{u}[i] + v[i] \,\big)
$$

proves claim (i).

(ii). Using the fact that $\|A(\hat{u}+v)-d\|^2 = \|A\hat{u}-d\|^2 + \|Av\|^2 + 2\langle Av,\, A\hat{u}-d\rangle$, one obtains

$$
\begin{aligned}
v \in \mathrm{B}_\infty(0,\rho)\setminus \mathrm{K}_{\hat{\sigma}} \quad \Rightarrow \quad & \mathcal{F}_d(\hat{u}+v) = \|A\hat{u}-d\|^2 + \|Av\|^2 + 2\langle Av,\, A\hat{u}-d\rangle \\
\Big[\text{by Lemma 2.1(i)}\Big] \qquad & \qquad +\beta\sum_{i\in\hat{\sigma}}\phi\left(\hat{u}[i]\right) + \beta\sum_{i\in\hat{\sigma}^c}\phi\left(v[i]\right) \\
\Big[\text{using (3)}\Big] \qquad & = \mathcal{F}_d(\hat{u}) + \|Av\|^2 + 2\langle Av,\, A\hat{u}-d\rangle + \beta\sum_{i\in\hat{\sigma}^c}\phi\left(v[i]\right) \\
& \geqslant \mathcal{F}_d(\hat{u}) - \Big|2\langle v,\, A^T(A\hat{u}-d)\rangle\Big| + \beta\|v_{\hat{\sigma}^c}\|_0 \\
\Big[\text{by Hölder's inequality}\Big] \qquad & \geqslant \mathcal{F}_d(\hat{u}) - 2\|v\|_\infty\, \|A^T(A\hat{u}-d)\|_1 + \beta\|v_{\hat{\sigma}^c}\|_0\ .
\end{aligned}
\tag{65}
$$

If $\sharp\,\hat{\sigma}^c = 0$, then $\mathrm{K}_{\hat{\sigma}} = \mathbb{R}^{\mathsf{N}}$, so $v\in\mathbb{R}^0$ and $\|v\|_\infty = 0$; hence we have the inequality.

Let $\sharp\,\hat{\sigma}^c \geqslant 1$. For $v \notin \mathrm{K}_{\hat{\sigma}}$, there at least one index $i\in\hat{\sigma}^c$ such that $v[i]\neq 0$; hence $\|v_{\hat{\sigma}^c}\|_0 \geqslant 1$. The definition of ρ in (19) shows that

$$
\begin{aligned}
v \in \mathrm{B}_\infty(0,\rho)\setminus \mathrm{K}_{\hat{\sigma}} \ \Rightarrow\ & -\|v\|_\infty > -\rho \geqslant -\frac{\beta}{2\Big(\|A^T(A\hat{u}-d)\|_1 + 1\Big)} \\
\Rightarrow\ & -2\|v\|_\infty\,\|A^T(A\hat{u}-d)\|_1 + \beta\|v_{\hat{\sigma}^c}\|_0 > -\frac{2\beta\|A^T(A\hat{u}-d)\|_1}{2\Big(\|A^T(A\hat{u}-d)\|_1 + 1\Big)} + \beta > 0\ .
\end{aligned}
$$

Introducing the last inequality into (65) shows that for $\sharp\,\hat{\sigma}^c \geqslant 1$, the inequality in (ii) is strict.

8.2. Proof of Proposition 4.1. If $\hat{u}=0$, the statement is obvious. We focus on $\hat{u}\neq 0$. For an arbitrary $i\in\mathbb{I}_{\mathsf{N}}$, define

$$
\hat{u}^{(i)} \overset{\text{def}}{=} \big(\hat{u}[1],\cdots,\hat{u}[i-1],\, 0,\, \hat{u}[i+1],\cdots,\hat{u}[\mathsf{N}]\big) \in \mathbb{R}^{\mathsf{N}}\ .
$$

We shall use the equivalent formulation of $\mathcal{F}_d$ given in (3). Clearly[25],

$$
\mathcal{F}_d(\hat{u}) = \mathcal{F}_d\big(\hat{u}^{(i)} + e_i\hat{u}[i]\big) = \|A\hat{u}^{(i)} + a_i\hat{u}[i] - d\|^2 + \beta\sum_{j\in\mathbb{I}_{\mathsf{N}}}\phi\left(\hat{u}^{(i)}[j]\right) + \phi\left(\hat{u}[i]\right)\ .
$$

Consider $f:\mathbb{R}\to\mathbb{R}$ as given below

$$
f(t) \overset{\text{def}}{=} \mathcal{F}_d\left(\hat{u}^{(i)} + e_i t\right)\ .
\tag{66}
$$

Since $\hat{u}$ is a global minimizer of $\mathcal{F}_d$, for any $i\in\mathbb{I}_{\mathsf{N}}$, we have

$$
\begin{aligned}
f\left(\hat{u}[i]\right) &= \mathcal{F}_d\big(\hat{u}^{(i)} + e_i\hat{u}[i]\big) \\
&\leqslant \mathcal{F}_d\big(\hat{u}^{(i)} + e_i\, t\big) = f(t) \quad \forall\, t\in\mathbb{R}\ .
\end{aligned}
$$

[25] Using the definition of $\hat{u}^{(i)}$, we have $\hat{u}^{(i)} = A_{(\mathbb{I}_{\mathsf{N}}\setminus\{i\})}\hat{u}_{(\mathbb{I}_{\mathsf{N}}\setminus\{i\})}$, hence $A\hat{u}^{(i)}$ is independent of $\hat{u}[i]$.

Equivalently, for any $i \in \mathbb{I}_\mathsf{N}$, $f(\hat{u}[i])$ is the global minimum of $f(t)$ on $\mathbb{R}$. Below we will determine the global minimizer(s) $\hat{t} = \hat{u}[i]$ of f as given in (66), i.e.,

$$\hat{t} = \hat{u}[i] = \arg\min_{t\in\mathbb{R}} f(t) \ .$$

In detail, the function f reads as

$$\begin{aligned} f(t) &= \|A\hat{u}^{(i)} + a_i t - d\|^2 + \beta \sum_{j\in\mathbb{I}_\mathsf{N}} \phi\left(\hat{u}^{(i)}[j]\right) + \beta\phi(t) \\ &= \|A\hat{u}^{(i)} - d\|^2 + \|a_i\|^2 t^2 + 2t\langle a_i, A\hat{u}^{(i)} - d\rangle + \beta \sum_{j\in\mathbb{I}_\mathsf{N}} \phi\left(\hat{u}^{(i)}[j]\right) + \beta\phi(t) \\ &= \|a_i\|^2 t^2 + 2t\langle a_i, A\hat{u}^{(i)} - d\rangle + \beta\phi(t) + C \ , \end{aligned} \tag{67}$$

where

$$C = \|A\hat{u}^{(i)} - d\|^2 + \beta \sum_{j\in\mathbb{I}_\mathsf{N}} \phi\left(\hat{u}^{(i)}[j]\right) \ .$$

Note that C does not depend on t. The function f has two local minimizers denoted by $\hat{t}_0$ and $\hat{t}_1$. The first is

$$\hat{t}_0 = 0 \quad \Rightarrow \quad f(\hat{t}_0) = C \ . \tag{68}$$

The other one, $\hat{t}_1 \neq 0$, corresponds to $\phi(t) = 1$. From (67), $\hat{t}_1$ solves

$$2\|a_i\|^2 t + 2\langle a_i, A\hat{u}^{(i)} - d\rangle = 0 \ .$$

Recalling that $a_i \neq 0$, $\forall\, i \in \mathbb{I}_\mathsf{N}$ (see (8)), it follows that

$$\hat{t}_1 = -\frac{\langle a_i, A\hat{u}^{(i)} - d\rangle}{\|a_i\|^2} \quad \Rightarrow \quad f(\hat{t}_1) = -\frac{\langle a_i, A\hat{u}^{(i)} - d\rangle^2}{\|a_i\|^2} + \beta + C \ . \tag{69}$$

Next we check whether $\hat{t}_0$ or $\hat{t}_1$ is a global minimizer of f. From (68) and (69) we get

$$f(\hat{t}_0) - f(\hat{t}_1) = \frac{\langle a_i, A\hat{u}^{(i)} - d\rangle^2}{\|a_i\|^2} - \beta \ .$$

Furthermore,

$$\begin{aligned} f(\hat{t}_0) < f(\hat{t}_1) \quad &\Rightarrow \quad \hat{u}[i] = \hat{t}_0 = 0 \ , \\ f(\hat{t}_1) < f(\hat{t}_0) \quad &\Rightarrow \quad \hat{u}[i] = \hat{t}_1 = -\frac{\langle a_i, A\hat{u}^{(i)} - d\rangle}{\|a_i\|^2} \ , \\ f(\hat{t}_0) = f(\hat{t}_1) \quad &\Rightarrow \quad \hat{t}_0 \text{ and } \hat{t}_1 \text{ are global minimizers of } f. \end{aligned} \tag{70}$$

In particular, we have

$$\begin{aligned} & f(\hat{t}_1) \leqslant f(\hat{t}_0) \quad \Leftrightarrow \quad \langle a_i, A\hat{u}^{(i)} - d\rangle^2 \geqslant \beta\|a_i\|^2 && (71) \\ & \left[\text{by } (70)\right] \qquad \Rightarrow \ |\hat{u}[i]| = \frac{\left|\langle a_i, A\hat{u}^{(i)} - d\rangle\right|}{\|a_i\|^2} \\ & \left[\text{by } (71)\right] \qquad\qquad \geqslant \frac{\sqrt{\beta}\|a_i\|}{\|a_i\|^2} = \frac{\sqrt{\beta}}{\|a_i\|} \ . \end{aligned}$$

It is clear that the conclusion holds true for any $i \in \mathbb{I}_\mathsf{N}$.

8.3. Proof of Proposition 4.3. The asymptotic function $(\mathcal{F}_d)_\infty(v)$ of $\mathcal{F}_d$ can be calculated according to[26] [3, Theorem 2.5.1]

$$(\mathcal{F}_d)_\infty(v) = \liminf_{\substack{v' \to v \\ t \to \infty}} \frac{\mathcal{F}_d(tv')}{t} .$$

Then

$$\begin{aligned}(\mathcal{F}_d)_\infty(v) &= \liminf_{\substack{v' \to v \\ t \to \infty}} \frac{\|Av'-d\|^2+\beta\|v'\|_0}{t} \\ &= \liminf_{\substack{v' \to v \\ t \to \infty}} \left(t\|Av'\|^2 - 2\langle d, Av'\rangle + \frac{\|d\|^2+\beta\|v'\|_0}{t}\right) \\ &= \begin{cases} 0 & \text{if } \; v \in \ker(A) \ , \\ +\infty & \text{if } \; v \notin \ker(A) \ . \end{cases}\end{aligned}$$

Hence

$$\ker\left((\mathcal{F}_d)_\infty\right) = \ker(A) \ , \tag{72}$$

where $\ker\left((\mathcal{F}_d)_\infty\right) = \{v \in \mathbb{R}^{\mathsf{N}} \ : \ (\mathcal{F}_d)_\infty(v) = 0\}$.

Let $\{v_k\}$ satisfy (37) with $v_k\,\|v_k\|^{-1} \to \bar{v} \in \ker(A)$. Below we compare the numbers $\|v_k\|_0$ and $\|v_k - \rho\bar{v}\|_0$ where $\rho > 0$. There are two options.

1. Consider that $i \in \sigma(\bar{v})$, that is, $\bar{v}[i] = \lim_{k\to\infty} v_k[i]\,\|v_k\|^{-1} \neq 0$. Then $|\,v_k[i]\,| > 0$ for all but finitely many k as otherwise, $v_k[i]\,\|v_k\|^{-1}$ would converge to 0. Therefore, there exists k_i such that

$$|\,v_k[i] - \rho\,\bar{v}[i]\,| \geqslant 0 \quad \text{and} \quad |\,v_k[i]\,| > 0 \quad \forall\, k \geqslant k_i \ . \tag{73}$$

2. If $i \in (\sigma(\bar{v}))^c$, i.e. $\bar{v}[i] = 0$, then clearly

$$v_k[i] - \rho\,\bar{v}[i] = v_k[i] \ . \tag{74}$$

Combining (73) and (74), the definition of $\|\cdot\|_0$ using ϕ in (2) shows that

$$\|v_k - \rho\,\bar{v}\|_0 \leqslant \|v_k\|_0 \quad \forall\, k \geqslant k_0 \stackrel{\text{def}}{=} \max_{i\in\sigma(\bar{v})} k_i \ . \tag{75}$$

By (72), $A\bar{v} = 0$. This fact, jointly with (75), entails that

$$\begin{aligned}\mathcal{F}_d(v_k - \rho\bar{v}) &= \|A(v_k + \rho\bar{v}) - d\|^2 + \beta\|v_k - \rho\bar{v}\|_0 \\ &= \|Av_k - d\|^2 + \beta\|v_k - \rho\bar{v}\|_0 \\ &\leqslant \|Av_k - d\|^2 + \beta\|v_k\|_0 = \mathcal{F}_d(v_k) \quad \forall\, k \geqslant k_0 \ .\end{aligned}$$

[26] In the nonconvex case, the notion of asymptotic functions and the representation formula were first given by J.P. Dedieu [12].

It follows that for any $k \geqslant k_0$ we have

$$v_k \in \text{lev}\,(\mathcal{F}_d, \lambda_k) \quad \Rightarrow \quad v_k - \rho\bar{v} \in \text{lev}\,(\mathcal{F}_d, \lambda_k)\ ,$$

and thus $\mathcal{F}_d$ satisfies Definition 4.2.

8.4. Proof of Proposition 4.5. Given $\mathsf{K} \in \mathbb{I}_{\mathsf{M}-1}$, set

$$\mathrm{U}_{\mathsf{K}+1} \stackrel{\text{def}}{=} \bigcup_{\omega\subset\mathbb{I}_{\mathsf{N}}} \{\bar{u} \ : \ \bar{u} \ \text{ solves } (\,\mathcal{P}_\omega\,) \quad \text{and} \quad \|\bar{u}\|_0 \geqslant \mathsf{K}+1\}\ . \tag{76}$$

- Let $\mathrm{U}_{\mathsf{K}+1} \neq \varnothing$. By Proposition 2.3, for any $\beta > 0$, $\mathcal{F}_d$ has a (local) minimum at each $\bar{u} \in \mathrm{U}_{\mathsf{K}+1}$. Thus

$$\bar{u} \ \text{ is a (local) minimizer of } \mathcal{F}_d \text{ and } \ \|\bar{u}\|_0 \geqslant \mathsf{K}+1 \quad \Leftrightarrow \quad \bar{u} \in \mathrm{U}_{\mathsf{K}+1}\ . \tag{77}$$

  Then for any $\beta > 0$

$$\mathcal{F}_d(\bar{u}) \geqslant \beta(\mathsf{K}+1) \quad \forall\, \bar{u} \in \mathrm{U}_{\mathsf{K}+1}\ . \tag{78}$$

  Let $\widetilde{u}$ be defined by[27]:

$$\widetilde{u} \ \text{ solves } (\,\mathcal{P}_\omega\,) \ \text{ for some } \ \omega \in \Omega_{\mathsf{K}}\ .$$

  Then

$$\|\widetilde{u}\|_0 \leqslant \mathsf{K}\ . \tag{79}$$

  Set β and β_{K} according to

$$\beta > \beta_{\mathsf{K}} \stackrel{\text{def}}{=} \|A\widetilde{u} - d\|^2\ . \tag{80}$$

  For such a β we have

$$\begin{aligned} \mathcal{F}_d(\widetilde{u}) &= \|A\widetilde{u} - d\|^2 + \beta\|\widetilde{u}\|_0 \\ \Big[\text{by (79) and (80)}\Big] &< \beta + \beta\mathsf{K} = \beta(\mathsf{K}+1) \\ \Big[\text{by (78)}\Big] &\leqslant \mathcal{F}_d(\bar{u}) \quad \forall\, \bar{u} \in \mathrm{U}_{\mathsf{K}+1}\ . \end{aligned}$$

  Let $\hat{u}$ be a global minimizer of $\mathcal{F}_d$. Then

$$\mathcal{F}_d(\hat{u}) \leqslant \mathcal{F}_d(\widetilde{u}) < \mathcal{F}_d(\bar{u}) \quad \forall\, \bar{u} \in \mathrm{U}_{\mathsf{K}+1}\ .$$

  Using (76)-(77), we find

$$\|\hat{u}\|_0 \leqslant \mathsf{K}\ .$$

- $\mathrm{U}_{\mathsf{K}+1} = \varnothing$ entails that[28]

$$\bar{u} \ \text{ solves } (\,\mathcal{P}_\omega\,) \ \text{ for } \ \omega \subset \mathbb{I}_{\mathsf{N}}, \ \ \sharp\,\omega \geqslant \mathsf{K}+1 \quad \Rightarrow \quad \|\bar{u}\|_0 \leqslant \mathsf{K}\ . \tag{81}$$

  Let $\hat{u}$ be a global minimizer of $\mathcal{F}_d$. By (81) we have

$$\|\hat{u}\|_0 \leqslant \mathsf{K}\ .$$

According to Theorem 4.4(ii), any global minimizer of $\mathcal{F}_d$ is strict, hence $\sigma(\hat{u}) \in \overline{\Omega}_{\mathsf{K}}$.

[27] Such a $\widetilde{u}$ always exists; see subsection 1.1. By Proposition 2.3 and Theorem 3.2, it is uniquely defined.
[28] Let $A = (e_1, e_2, e_3, e_4, e_1) \in \mathbb{R}^{4\times 5}$ and $d = e_1 \in \mathbb{R}^4$. For $\mathsf{K} = \mathsf{M} - 1 = 3$ one can check that $\mathrm{U}_{\mathsf{K}+1} = \varnothing$.

Acknowledgments. The author would like to thank the anonymous Reviewer 2 for very helpful remarks and suggestions.

REFERENCES

[1] H. Attouch, J. Bolte, and F. Svaiter, *Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms, forwardbackward splitting, and regularized gaussseidel methods*, Mathematical Programming, 137 (2013), pp. 91–129.

[2] A. Auslender, *Existence of optimal solutions and duality results under weak conditions*, Mathematical Programming, 88 (2000), pp 45–59.

[3] A. Auslender and M. Teboulle, *Asymptotic Cones and Functions in Optimization and Variational Inequalities*, Springer, New York, 2003.

[4] C. Baiocchi, G. Buttazo, F. Gastaldi, and F. Tomarelli, *General existence theorems for unilateral problems in continuum mechanics*, Arch. Rational Mech. Anal., 100 (1998), pp. 149–189.

[5] J. E. Besag, *On the statistical analysis of dirty pictures (with discussion)*, Journal of the Royal Statistical Society B, 48 (1986), pp. 259–302.

[6] ———, *Digital image processing: Towards Bayesian image analysis*, Journal of Applied Statistics, 16 (1989), pp. 395–407.

[7] T. Blumensath and M. Davies, *Iterative thresholding for sparse approximations*, Journal of Fourier Analysis and Applications, 14 (2008), pp. 629–654.

[8] A. M. Bruckstein, D. L. Donoho, and M. Elad, *From sparse solutions of systems of equations to sparse modeling of signals and images*, SIAM Review, 51 (2009), pp. 34–81.

[9] E. Chouzenoux, A. Jezierska, J.-C. Pesquet, and H. Talbot, *A majorize-minimize subspace approach for $\ell_2 - \ell_0$ image regularization*, SIAM J. Imaging Sci., 6 (2013), pp. 563–591.

[10] *Compressive Sensing Resources*, references and software, http://dsp.rice.edu/cs, 2013.

[11] G. Davis, S. Mallat, and M. Avellaneda, *Adaptive greedy approximations*, Constructive approximation, 13 (1997), pp. 57–98.

[12] J. Dedieu, *Cône asymptote d'un ensemble non convexe. Application à l'optimisation.*, Compte-rendus de l'académie des sciences, 287 (1977), pp. 91–103.

[13] G. Demoment, *Image reconstruction and restoration : Overview of common estimation structure and problems*, IEEE Transactions on Acoustics Speech and Signal Processing, ASSP-37 (1989), pp. 2024–2036.

[14] B. Dong and Y. Zhang, *An efficient algorithm for ℓ_0 minimization in wavelet frame based image restoration*,J. Sci. Comput., 54 (2013), pp. 350-368.

[15] D. L. Donoho and I. M. Johnstone, *Ideal spatial adaptation by wavelet shrinkage*, Biometrika, 81 (1994), pp. 425–455.

[16] L. C. Evans and R. F. Gariepy, *Measure theory and fine properties of functions*, Studies in Advanced Mathematics, CRC Press, Roca Baton, FL, 1992.

[17] J. Fan and R. Li, *Statistical challenges with high dimensionality: feature selection in knowledge discovery*, in Proceedings of the International Congress of Mathematicians, Madrid, Spain, vol. 3, European Mathematical Society, 2006, pp. 595–622.

[18] M. Fornasier and R. Ward, *Iterative thresholding meets free-discontinuity problems*, Foundations of Computational Mathematics, 10 (2010), pp. 527–567.

[19] G. Gasso, A. Rakotomamonjy, and S. Canu, *Recovering sparse signals with a certain family of nonconvex penalties and DC programming*, IEEE Transactions on Signal Processing, 57 (2009), pp. 4686–4698.

[20] S. Geman and D. Geman, *Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images*, IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI-6 (1984), pp. 721–741.

[21] G. Golub and C. Van Loan, *Matrix Computations*, The Johns Hopkins University Press, Baltimore and London, 3 ed., 1996.

[22] J. Haupt and R. Nowak, *Signal reconstruction from noisy random projections*, IEEE Transactions on Information Theory, 52 (2006), pp. 4036–4048.

[23] Y. Leclerc, *Constructing simple stable descriptions for image partitioning*, International Journal of Computer Vision, 3 (1989), pp. 73–102.

[24] S. Li, *Markov Random Field Modeling in Computer Vision*, Springer-Verlag, London, UK, 1 ed., 1995.

[25] Y. Liu and Y. Wu, *Variable selection via a combination of the l_0 and l_1 penalties*, Journal of Computational and Graphical Statistics, 16 (2007), pp. 782–798.

[26] Z. Lu and Y. Zhang, *Sparse approximation via penalty decomposition methods*, preprint, arXiv:1205.2334v2 [cs.LG], 2012.

[27] J. Lv and Y. Fan, *A unified approach to model selection and sparse recovery using regularized least squares*, The Annals of Statistics, 37 (2009), pp. 3498–3528.

[28] S. Mallat, *A Wavelet Tour of Signal Processing (The sparse way)*, Academic Press, London, 3 ed., 2008.

[29] C. D. Meyer, *Matrix Analysis and Applied Linear Algebra*, SIAM, 2000.

[30] A. J. Miller, *Subset Selection in Regression*, Chapman and Hall, London, U.K., 2 ed., 2002.

[31] J. Neumann, C. Schörr, and G. Steidl, *Combined SVM-based Feature Selection and classification*, Machine Learning 61, 2005, pp. 129–150.

[32] M. Nikolova, *Analysis of the recovery of edges in images and signals by minimizing nonconvex regularized least-squares*, SIAM Journal on Multiscale Modeling and Simulation, 4 (2005), pp. 960–991.

[33] M. Robini, A. Lachal, and I. Magnin, *A stochastic continuation approach to piecewise constant reconstruction*, IEEE Transactions on Image Processing, 16 (2007), pp. 2576–2589.

[34] M. C. Robini and I. E. Magnin, *Optimization by stochastic continuation*, SIAM Journal on Imaging Sciences, 3 (2010), pp. 1096–1121.

[35] M. C. Robini and P.-J. Reissman, *From simulated annealing to stochastic continuation: a new trend in combinatorial optimization*, Journal of Global Optimization, 56 (2013), pp. 185–215.

[36] W. Rudin, *Principles of Mathematical analysis*, Mathematics Series, McGraw-Hill, Berlin, 1976.

[37] ———, *Real and Complex Analysis*, Mathematics Series, McGraw-Hill, Berlin, 1987.

[38] E. M. Stein and R. Shakarchi, *Real Analysis: Measure Theory, Integration, and Hilbert Spaces*, Princeton University Press, 2005.

[39] M. Thiao, T. P. Dinh, and A. L. Thi, *DC Programming Approach for a Class of Nonconvex Programs Involving l0 Norm*, Communications in Computer and Information Science, vol. 14, Springer, 2008, pp. 348–357.

[40] J. Tropp, *Just relax: convex programming methods for identifying sparse signals in noise*, IEEE Transactions on Information Theory, 52 (2006), pp. 1030–1051.

[41] M. Vetterli and Kovačević, *Wavelets and subband coding*, Prentice Hall PTR, 1995.

[42] Y. Zhang, B. Dong, and Z. Lu, *ℓ_0 minimization of wavelet frame based image restoration*, Mathematics of Computation, 82 (2013), pp. 995-1015.